THE MATHEMATICAL STATISTICIAN

Vol. 25, No. 4

ON THE CONSTRUCTION OF GROUP DIVISIBLE INCOMPLETE BLOCK DESIGNS[1]

By R. C. Bose, S. S. Shrikhande and K. N. Bhattacharya

University of North Carolina, Nagpur College of Science and University of Kansas, and University of Calcutta

1. Summary. It has been shown in [1] that all partially balanced incomplete block (PBIB) designs with two associate classes, can be divided into a small number of types according to the nature of the association relations among the treatments. One simple and important type is the group divisible (GD). The combinatorial properties of GD designs have been studied in [2] and the analysis along with that for other types is given in [1]. Here we give methods of constructing GD designs. These designs are likely to prove useful in agricultural, genetic and industrial experiments.

2. Introduction. An incomplete block design with v treatments each replicated r times in b blocks of size k is said to be group divisible (GD) if the treatments can be divided into m groups, each with n treatments, so that the treatments belonging to the same group occur together in λ_1 blocks and treatments belonging to different groups occur together in λ_2 blocks. If $\lambda_1 = \lambda_2 = \lambda$ (say) then every pair of treatments occurs together in λ blocks and the design reduces to the well known balanced incomplete block (BIB) design.

It has been shown in [2] that the parameters v, b, r, k, m, n, λ_1 and λ_2 satisfy the following relations and inequalities.

$$v = mn, \qquad bk = vr \tag{2.0}$$

$$\lambda_1(n-1) + \lambda_2 n(m-1) = r(k-1) \tag{2.1}$$

$$Q = r - \lambda_1 \geqq 0, \qquad P = rk - v\lambda_2 \geqq 0. \tag{2.2}$$

The GD designs were divided into three classes: (a) Singular GD designs characterized by $Q = 0$, (b) Semi-regular GD designs characterized by $Q > 0$, $P = 0$, (c) Regular GD designs characterized by $Q > 0$, $P > 0$. The combinatorial properties of each class were separately studied. These will be referred to at appropriate places so far as they are relevant to the problem of construction of GD designs, which is the main concern of this paper. We shall confine ourselves to the practically useful range $v \geqq 10$, $r \leqq 10$, $k \leqq 10$, and choose λ_1 and λ_2 not to exceed 3, except for a few singular and semi-regular designs of special interest.

As noted in [1] GD designs besides being a sub-class of PBIB designs [3], [4] with two associate classes, can also be regarded as a sub-class of inter- and intra-group balanced incomplete block (IIGBI) designs [5].

Received 10/14/52.

[1] This research was supported in part by the United States Air Force under Contract AF18(600)-83 monitored by the Office of Scientific Research.

3. Some types of balanced incomplete block (BIB) designs.

(a) The construction of GD designs can in many instances be made to depend on known solutions for BIB designs [6], [7], [8], [9]. We shall here bring together certain results with a view to subsequent use. The parameters of a BIB design will be denoted by a starred letter in order to distinguish them from the parameters of GD designs. Thus the number of treatments will be denoted by v^*, the number of blocks by b^*, the number of replications of each treatment by r^*, the number of treatments in each block by k^*, and the number of times any two treatments occur together in a block by λ^*. The design is said to be resolvable [10] if the blocks can be grouped in such a way that each group contains a complete replication.

(b) The simplest type of BIB design is the unreduced type with $k = 2$, the blocks of which are obtained by taking all possible pairs of t treatments. The parameters are

$$(3.0) \quad v^* = t, \qquad b^* = t(t-1)/2, \qquad r^* = t-1, \qquad k^* = 2, \qquad \lambda^* = 1$$

We shall later use the fact that when t is even, the solution can be expressed in a resolvable form. For example, if $t = 6$, then we can write the 15 blocks as

$$(3.1) \qquad \begin{array}{lll} (1,4), & (2,3), & (0,\infty) \\ (2,0), & (3,4), & (1,\infty) \\ (3,1), & (4,0), & (2,\infty) \\ (4,2), & (0,1), & (3,\infty) \\ (0,3), & (1,2), & (4,\infty) \end{array}$$

where the treatments are 0, 1, 2, 3, 4 and ∞, and the three blocks in any particular row of (3.1) give a complete replication. In the general case when $t = 2u$ the solution can be generated by developing the initial blocks

$$(3.2) \quad (1, 2u-2), (2, 2u-3), \cdots, (u-1, u), (0, \infty) \mod (2u-1),$$

the treatment ∞ remaining unchanged. The designs (3.0) will be referred to as belonging to series (u).

(c) BIB designs with parameters

$$(3.3) \qquad v^* = s^2, \qquad b^* = s^2 + s, \qquad r^* = s+1, \qquad k^* = s, \quad \lambda^* = 1$$

may be said to belong to the orthogonal series 1 (0S1). They are also called balanced lattices [11], and can be obtained from a complete set of orthogonal Latin squares [6], [7]. They can, however, be more readily obtained by using certain difference sets [12] due to one of the authors, which have been given in Table I, and whose use is explained below.

For example let $s = 4$. If we develop the difference set for $s = 4$, mod $(s^2 - 1)$, we get fifteen blocks of the BIB design

$$(3.35) \qquad v^* = 16, \qquad b^* = 20, \qquad r^* = 5, \qquad k^* = 4, \qquad \lambda^* = 1.$$

They are given by the columns of the scheme

$$(3.4)\qquad \begin{array}{rrrrrrrrrrrrrrr} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 0 \\ 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 0 & 1 & 2 \\ 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 0 & 1 & 2 & 3 \\ 12 & 13 & 14 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11. \end{array}$$

The remaining blocks are obtained by starting with the block $0, s+1, 2(s+1), \infty$ and deriving other blocks by adding $1, 2, \cdots, s$ to the treatments of this block, remembering that ∞ is invariant under addition. Thus 5 other blocks are given by the columns of the scheme

$$(3.45)\qquad \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 & 9 \\ 10 & 11 & 12 & 13 & 14 \\ \infty & \infty & \infty & \infty & \infty. \end{array}$$

TABLE I

Difference sets for generating BIB designs belonging to the orthogonal series 0S 1

s	Difference set	Modulus
2	1, 2	mod (3)
3	1, 6, 7	mod (8)
4	1, 3, 4, 12	mod (15)
5	1, 3, 16, 17, 20	mod (24)
7	1, 2, 5, 11, 31, 36, 38	mod (48)
8	1, 6, 8, 14, 38, 48, 49, 52	mod (63)
9	1, 13, 35, 48, 49, 66, 72, 74, 77	mod (80)

The design is resolvable, the ith replication being obtained by taking the ith block from (3.45), and the ith and every succeeding $(s+1)$st block from (3.4). We may thus rearrange the twenty blocks and get the design in the form, where the replications are separated by vertical lines,

$$(3.5)\qquad \begin{array}{cccc|cccc|cccc|cccc|cccc} 1 & 6 & 11 & 0 & 2 & 7 & 12 & 1 & 3 & 8 & 13 & 2 & 4 & 9 & 14 & 3 & 5 & 10 & 0 & 4 \\ 3 & 8 & 13 & 5 & 4 & 9 & 14 & 6 & 5 & 10 & 0 & 7 & 6 & 11 & 1 & 8 & 7 & 12 & 2 & 9 \\ 4 & 9 & 14 & 10 & 5 & 10 & 0 & 11 & 6 & 11 & 1 & 12 & 7 & 12 & 2 & 13 & 8 & 13 & 3 & 14 \\ 12 & 2 & 7 & \infty & 13 & 3 & 8 & \infty & 14 & 4 & 9 & \infty & 0 & 5 & 10 & \infty & 1 & 6 & 11 & \infty. \end{array}$$

(d) BIB designs with parameters

$$(3.6)\qquad v^* = b^* = s^2 + s + 1, \qquad r^* = k^* = s + 1, \qquad \lambda^* = 1$$

may be said to belong to the orthogonal series 2 (0S 2). The solution for any design of 0S 2 can be obtained from the corresponding design of 0S 1 by taking

$s + 1$ new treatments, and by adding the ith new treatment to each block of the ith replication, and finally adding a new block containing all the new treatments. A solution is, however, more readily obtained by using the following difference sets due to Singer [13], which have been given in Table II.

Thus the blocks for the BIB design

$$v^* = b^* = 13, \quad r^* = k^* = 4, \quad \lambda^* = 1 \tag{3.65}$$

obtained by using the difference set corresponding to $s = 3$, are given by the columns of the scheme

$$\begin{array}{cccccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 0 \\ 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 0 & 1 & 2 \\ 9 & 10 & 11 & 12 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8. \end{array} \tag{3.7}$$

TABLE II

Difference sets for generating BIB designs belonging to the orthogonal series OS 2

s	Difference set	Modulus
2	0, 1, 3	mod (7)
3	0, 1, 3, 9	mod (13)
4	0, 1, 4, 14, 16	mod (21)
5	0, 1, 3, 8, 12, 18	mod (31)
7	0, 1, 3, 13, 32, 36, 43, 52	mod (57)
8	0, 1, 3, 7, 15, 31, 36, 54, 63	mod (73)
9	0, 1, 3, 9, 27, 49, 56, 61, 77, 81	mod (91)
11	0, 1, 3, 12, 20, 34, 38, 81, 88, 94, 104, 109	mod (133)

(e) BIB designs which are resolvable or (and) for which $\lambda = 1$ are especially important for the construction of GD designs. We present in Table III designs of the type for which $r^* \leqq 11$, and which do not belong to the series u, $0S$ 1 or $0S$ 2 already considered. The reference to the series is in the notation used in [8]. In each case the complete solution can be developed from certain initial blocks.

Designs marked + are resolvable. For (2), (5) and (9) the initial blocks provide a complete replication. Hence, in developing, the replications remain separate. For (6) the first seven blocks provide a complete replication and when developed yield replications I–VII. Each of the other three initial blocks when developed yields a complete replication. The solutions given here have been taken or adapted from [8], [14] and [15]. In developing the initial blocks the suffixes and ∞ should be kept invariant. (For the use of binary symbols see Section 6.)

TABLE III

BIB designs which are resolvable or (and) for which $\lambda = 1$.

Serial no.	Series	Parameters v^*	b^*	r^*	k^*	λ^*	Initial blocks	Modulus
(1)	T_2	13	26	6	3	1	(1,3,9), (2,6,5)	mod (13)
(2)+	T_1	15	35	7	3	1	$(1_1,2_1,4_1)$, $(3_1,1_2,5_2)$, $(6_1,2_2,3_2)$, $(5_1,4_2,6_2)$, $(0_1,0_2,\infty)$	mod (7)
(3)	F_1	25	50	8	4	1	(00,01,41,13), (00,32,21,02)	mod (5, 5)
(4)	T_2	19	57	9	3	1	(1,7,11), (2,14,3), (4,9,6)	mod (19)
(5)+	F_2	28	63	9	4	1	$(01_1,02_1,10_2,20_2)$, $(21_1,12_1,22_2,11_2)$, $(01_2,02_2,10_3,20_3)$, $(21_2,12_2,22_3,11_3)$, $(01_3,02_3,10_1,20_1)$, $(21_3,12_3,22_1,11_1)$, $(00_1,00_2,00_3,\infty)$	mod (3, 3)
(6)+	(T_1)	21	70	10	3	1	$(0_1,0_2,0_3)$, $(1_1,2_1,4_1)$, $(1_2,2_2,4_2)$, $(1_3,2_3,4_3)$, $(3_1,5_2,6_3)$, $(3_2,5_3,6_1)$, $(3_3,5_1,6_2)$ Reps I–VII; $(1_1,2_3,4_2)$ Rep VIII; $(1_2,2_1,4_3)$ Rep IX; $(1_3,2_2,4_1)$ Rep X	mod (7)
(7)	(G_1)	41	82	10	5	1	(1,37,16,18,10), (8,9,5,21,39)	mod (41)
(8)	(G_2)	45	99	11	5	1	$(01_1,02_1,10_3,20_4,00_2)$, $(21_1,12_1,22_3,11_4,00_2)$, $(01_2,02_2,10_4,20_4,00_3)$, $(21_2,12_2,22_4,11_4,00_3)$, $(01_3,02_3,10_5,20_5,00_4)$, $(21_3,12_3,22_5,11_5,00_4)$, $(01_4,02_4,10_1,20_1,00_5)$, $(21_4,12_4,22_1,11_1,00_5)$, $(01_5,02_5,10_2,20_2,00_1)$, $(21_5,12_5,22_2,11_2,00_1)$, $(00_1,00_2,00_3,00_4,00_5)$	mod (3, 3)
(9)+	(B_1)	8	14	7	4	3	(0,1,2,4), (3,5,6,∞)	mod (7)

4. Construction of singular GD designs. It has been shown in [2] that if in a BIB design with parameters v^*, b^*, r^*, k^*, λ^* we replace each treatment by a group of n treatments, we get a singular GD design with parameters

$$(4.0) \qquad \begin{array}{llll} v = nv^*, & b = b^*, & r = r^*, & k = nk^*, \\ m = v^*, & n = n, & \lambda_1 = r^*, & \lambda_2 = \lambda^*. \end{array}$$

Conversely, every singular GD design is obtainable in this way from a corresponding BIB design. The problem of constructing singular GD designs, there-

fore, offers no difficulty. However, if r^* and λ^* differ too much, then in the derived GD design, the accuracy of the within group and between group comparisons will appreciably differ. We give in Table IV some cases of practical interest.

A singular GD design may be considered to belong to the same series as the corresponding BIB design. The series has been shown along with the serial number in Table IV. It is clear that if a BIB design is resolvable the same is true of a GD design derived from it. Resolvability has been denoted by +.

As an example consider design (11) of Table IV. The blocks of the corresponding BIB design are given by (3.7). Replacing each treatment i by two treatments

TABLE IV

Parameters of some singular GD designs, and the corresponding BIB designs from which they are derivable

Serial no. and series	Parameters of BIB design					Parameters of corresponding singular GD design							
	v^*	b^*	r^*	k^*	λ^*	v	b	r	k	m	n	λ_1	λ_2
(1) $u+$	4	6	3	2	1	12	6	3	6	6	3	3	1
(2) $u+$	4	6	3	2	1	16	6	3	8	6	4	3	1
(3) $u+$	4	6	3	2	1	20	6	3	10	6	5	3	1
(4) $0S\ 2$	7	7	3	3	1	14	7	3	6	7	2	3	1
(5) $0S\ 2$	7	7	3	3	1	21	7	3	9	7	3	3	1
(6) u	5	10	4	2	1	10	10	4	4	5	2	4	1
(7) u	5	10	4	2	1	15	10	4	6	5	3	4	1
(8) u	5	10	4	2	1	20	10	4	8	5	4	4	1
(9) $0S\ 1+$	9	12	4	3	1	18	12	4	6	9	2	4	1
(10) $0S\ 1+$	9	12	4	3	1	27	12	4	9	9	3	4	1
(11) $0S\ 2$	13	13	4	4	1	26	13	4	8	13	2	4	1

i_1 and i_2, we see that the blocks of the GD design under consideration, are given by the columns of the scheme

$$
\begin{array}{cccccccccccc}
0_1 & 1_1 & 2_1 & 3_1 & 4_1 & 5_1 & 6_1 & 7_1 & 8_1 & 9_1 & 10_1 & 11_1 & 12_1 \\
0_2 & 1_2 & 2_2 & 3_2 & 4_2 & 5_2 & 6_2 & 7_2 & 8_2 & 9_2 & 10_2 & 11_2 & 12_2 \\
1_1 & 2_1 & 3_1 & 4_1 & 5_1 & 6_1 & 7_1 & 8_1 & 9_1 & 10_1 & 11_1 & 12_1 & 0_1 \\
1_2 & 2_2 & 3_2 & 4_2 & 5_2 & 6_2 & 7_2 & 8_2 & 9_2 & 10_2 & 11_2 & 12_2 & 0_2 \\
3_1 & 4_1 & 5_1 & 6_1 & 7_1 & 8_1 & 9_1 & 10_1 & 11_1 & 12_1 & 0_1 & 1_1 & 2_1 \\
3_2 & 4_2 & 5_2 & 6_2 & 7_2 & 8_2 & 9_2 & 10_2 & 11_2 & 12_2 & 0_2 & 1_2 & 2_2 \\
9_1 & 10_1 & 11_1 & 12_1 & 0_1 & 1_1 & 2_1 & 3_1 & 4_1 & 5_1 & 6_1 & 7_1 & 8_1 \\
9_2 & 10_2 & 11_2 & 12_2 & 0_2 & 1_2 & 2_2 & 3_2 & 4_2 & 5_2 & 6_2 & 7_2 & 8_2 .
\end{array} \tag{4.1}
$$

The treatments i_1 and i_2 belong to the same group ($i = 0, 1, \cdots, 12$). They occur together in the same block four times. Two treatments not belonging to the same group occur together in a block just once.

5. Method of "omitting varieties" for the generation of GD designs. Consider a BIB design with parameters

$$v^*, \quad b^*, \quad r^*, \quad k^*, \quad \lambda^* = 1. \tag{5.0}$$

A particular treatment θ occurs in r blocks. The remaining $v^* - 1 = r^*(k^* - 1)$ treatments can be divided into r^* groups, each containing $k^* - 1$ treatments, two treatments belonging to the same group if they occur together in the same block with θ. If we form a new design by omitting the treatment θ, and all the blocks containing it, we evidently get a GD design with parameters

$$\begin{array}{llll} v = v^* - 1, & b = b^* - r^*, & r = r^* - 1, & k = k^*, \\ m = r^*, & n = k^* - 1, & \lambda_1 = 0, & \lambda_2 = 1. \end{array} \tag{5.1}$$

THEOREM 1. *By omitting a particular treatment θ from a BIB design with parameters* (5.0), *we obtain a GD design with parameters* (5.1). *Two treatments belong to the same group if they occur together in the same block as θ.*

In particular, if we start with a BIB design belonging to the orthogonal series 0*S* 1, with parameters given by (3.3), we get a series of regular GD designs with parameters

$$v = b = s^2 - 1, \quad r = k = s, \quad m = s + 1, \quad n = s - 1, \quad \lambda_1 = 0, \quad \lambda_2 = 1. \tag{5.2}$$

The method of obtaining the blocks of a design of 0*S* 1 using the difference sets in Table I has already been explained. To get the corresponding design of (5.2), it is convenient to omit the treatment ∞. Thus taking $s = 4$, the blocks of the GD design

$$v = b = 15, \quad r = k = 4, \quad m = 5, \quad n = 3, \quad \lambda_1 = 1, \quad \lambda_2 = 0 \tag{5.3}$$

are given by the columns of the scheme (3.4), and the groups are given by the columns of (3.45), if the last row containing only ∞ is omitted.

The BIB designs (1)–(8) of Table III may also be employed to generate corresponding GD designs. For example the blocks of

$$v^* = 13, \quad b^* = 26, \quad r^* = 6, \quad k^* = 3, \quad \lambda^* = 1 \tag{5.4}$$

obtained by developing the initial blocks given in Table III are

$$\begin{array}{l} 1\ 2\ 3\ 4\ 5\ 6\ 7\ \ 8\ \ 9\ 10\ 11\ 12\ 0\ 2\ 3\ 4\ 5\ \ 6\ \ 7\ \ 8\ \ 9\ 10\ 11\ 12\ 0\ 1 \\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ \ 0\ \ 1\ 2\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ \ 0\ \ 1\ \ 2\ \ 3\ 4\ 5 \\ 9\ 10\ 11\ 12\ 0\ 1\ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7\ 8\ 5\ 6\ 7\ 8\ \ 9\ 10\ 11\ 12\ \ 0\ \ 1\ \ 2\ 3\ 4. \end{array} \tag{5.5}$$

Omitting the treatment 0, the blocks of the GD design

$$\begin{array}{llll} v = 12, & b = 20, & r = 5, & k = 3, \\ m = 6, & n = 2, & \lambda_1 = 0, & \lambda_2 = 1 \end{array} \tag{5.6}$$

are given by the columns of the scheme

$$\begin{array}{l} 1\ \ 2\ \ 3\ \ 4\ \ 6\ 7\ \ 8\ \ 9\ 10\ 12\ 2\ 3\ 4\ 5\ \ 6\ \ 7\ \ 8\ 11\ 12\ 1 \\ 3\ \ 4\ \ 5\ \ 6\ \ 8\ 9\ 10\ 11\ 12\ \ 1\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ \ 2\ \ 3\ 5 \\ 9\ 10\ 11\ 12\ \ 1\ 2\ \ 3\ \ 4\ \ 5\ \ 7\ 5\ 6\ 7\ 8\ \ 9\ 10\ 11\ \ 1\ \ 2\ 4\ , \end{array} \tag{5.7}$$

and the groups are given by the columns in

$$\begin{matrix} 5 & 11 & 2 & 9 & 10 & 4 \\ 7 & 6 & 8 & 12 & 1 & 3. \end{matrix} \tag{5.8}$$

We give in Table V parameters of BIB designs with $\lambda = 1$, together with the parameters of GD designs derivable from them by omitting a variety. Designs of the orthogonal series $0S\ 2$, and the semi-regular GD designs derivable from them have not been included, as the latter will be obtained in Section 7, as members of a more general class.

TABLE V

Parameters of BIB designs with $\lambda = 1$ not belonging to the series $0S\ 2$ and GD designs derivable from them by "omitting varieties"

Serial no. and series	Parameters of BIB design					Parameters of GD design							
	v^*	b^*	r^*	k^*	λ^*	v	b	r	k	m	n	λ_1	λ_2
(1) $0S\ 1$	16	20	5	4	1	15	15	4	4	5	3	0	1
(2) $0S\ 1$	25	30	6	5	1	24	24	5	5	6	4	0	1
(3) $0S\ 1$	49	56	8	7	1	48	48	7	7	8	6	0	1
(4) $0S\ 1$	64	72	9	8	1	63	63	8	8	9	7	0	1
(5) $0S\ 1$	81	90	10	9	1	80	80	9	9	10	8	0	1
(6) T_2	13	26	6	3	1	12	20	5	3	6	2	0	1
(7) T_1	15	35	7	3	1	14	28	6	3	7	2	0	1
(8) F_1	25	50	8	4	1	24	42	7	4	8	3	0	1
(9) T_2	19	57	9	3	1	18	48	8	3	9	2	0	1
(10) F_2	28	63	9	4	1	27	54	8	4	9	3	0	1
(11) T_1	21	70	10	3	1	20	60	9	3	10	2	0	1
(12) G_1	41	82	10	5	1	40	72	9	5	10	4	0	1
(13) G_2	45	99	11	5	1	44	88	10	5	11	4	0	1

6. Method of differences for generating GD designs.

(a) The method of differences has been extensively used in [8] and [9] for the construction of BIB designs. We shall here adapt it to the construction of GD designs. Consider a module M with a finite number of elements. To each element let there correspond h treatments, the treatments corresponding to the element x being

$$x_1, x_2, \cdots, x_h. \tag{6.0}$$

Thus there are $v = gh$ treatments. Treatments denoted by symbols with the same lower suffix i may be said to belong to the ith class.

Let $x_i^{(u)}$ and $x_j^{(v)}$ be two different treatments of the ith and jth classes respectively, where $x^{(u)}$ and $x^{(v)}$ are elements of M. Let

$$x^{(u)} - x^{(v)} = d, \qquad x^{(v)} - x^{(u)} = -d. \tag{6.1}$$

We then say that the pair of treatments $x_i^{(u)}$ and $x_j^{(v)}$ give rise to the difference d of the type $[i, j]$ and difference $-d$ of the type $[j, i]$. When $i = j$ the differences are called "pure" and when $i \neq j$ the differences are called "mixed". The differences d of the type $[i, j]$ and $-d$ of type $[j, i]$ are said to be "complementary" to one another. Thus every pair of treatments gives rise to a pair of complementary differences, one difference corresponding to each order of writing the treatments. Clearly there are h different types of pure differences and $h(h-1)$ different types of mixed differences. Since every nonzero element of M can appear in a pure difference, and every element (zero or nonzero) in a mixed difference, the total number of different possible differences is

$$h(g-1) + h(h-1)g = v(v-1)/g. \tag{6.15}$$

If θ is an arbitrary element of M and

$$x^{(\alpha)} = x^{(u)} + \theta, \qquad x^{(\beta)} = x^{(v)} + \theta, \tag{6.2}$$

then the pair of treatments $x_i^{(\alpha)}$ and $x_j^{(\beta)}$ give rise to the same pair of complementary differences as $x_i^{(u)}$ and $x_j^{(v)}$. Since θ can take g different values, we get g pairs of treatments giving rise to differences d and $-d$ of types $[i, j]$ and $[j, i]$ respectively, and it is easy to see that there are no other treatment pairs which give rise to the same differences. The $v(v-1)/2$ treatment pairs thus give rise to just $v(v-1)/g$ differences, which checks with (6.15).

Given an initial block B containing k treatments we can get g blocks by developing it in the following manner. Let θ be any arbitrary element of M. Then we get a new block B_θ corresponding to θ by replacing each treatment x_i in the initial block by x_i' where $x' = x + \theta$. By varying θ we get all the g required blocks. The initial block B gives rise to $k(k-1)$ differences namely, the differences which arise from the $k(k-1)/2$ pairs of treatments which can be formed from the treatments in B. If any pair of treatments occurs in B, then all the g pairs of treatments which give rise to the same differences as the given pair, occur in the corresponding positions in the blocks developed from B.

(b) THEOREM 2. *Let M be a module with m elements and to each element of M let there correspond n treatments. Let it be possible to find t initial blocks*

$$B_1, B_2, \cdots, B_t$$

each containing k treatments, and an initial group G containing n treatments such that

(i) *the $n(n-1)$ differences arising from G are all different, and*

(ii) *among the $k(k-1)t$ differences arising from the initial blocks each difference occurs λ_2 times, except those which arise from G, each of which occurs λ_1 times.*

Then by developing the initial blocks $B_1, B_2, \cdots, B_t$ we get the GD design with parameters $v = mn$, $b = mt$, $r = kt/n$, k, m, n, λ_1, λ_2, the group being obtained by developing the initial group G.

PROOF. Two treatments belong to the same group if and only if the differences arising from them occur among those arising from G. By the conditions of the theorem and what has been said before any such pair will occur among the developed blocks λ_1 times, and all other pairs will occur λ_2 times. Also in the developed blocks each treatment must occur in $(n-1)\lambda_1 + n(m-1)\lambda_2$ pairs. But if this treatment occurs in r blocks then this number of pairs is also $r(k-1)$. Hence r must be the same for all treatments and is given by

$$(n-1)\lambda_1 + n(m-1)\lambda_2 = r(k-1). \tag{6.25}$$

Again the total number of pairs in all the developed blocks is $mk(k-1)t/2$ and this must equal $mnr(k-1)/2$ since each treatment occurs in $r(k-1)$ pairs. Hence $r = kt/n$. This completes the proof.

In particular let M be the module of residue classes mod (m), and let the initial group G consist of treatments

$$0_1, 0_2, \cdots, 0_n. \tag{6.3}$$

Then to get a GD design with parameters $v, b, r, k, m, n, \lambda_1, \lambda_2$ we have to find t initial blocks such that among the $k(k-1)t$ differences arising from these blocks each pure difference and each nonzero mixed difference arises just λ_2 times, and each zero mixed difference arises λ_1 times. The designs (1)–(7) of Table VI have been obtained by using this special case of Theorem 2. For example for design (2) of Table VI, the complete set of blocks obtained by developing the given initial blocks mod (7), are given by the columns of the scheme

$$\begin{array}{cccccccccccccc} 1_1 & 2_1 & 3_1 & 4_1 & 5_1 & 6_1 & 0_1 & 1_2 & 2_2 & 3_2 & 4_2 & 5_2 & 6_2 & 0_2 \\ 2_1 & 3_1 & 4_1 & 5_1 & 6_1 & 0_1 & 1_1 & 2_2 & 3_2 & 4_2 & 5_2 & 6_2 & 0_2 & 1_2 \\ 4_1 & 5_1 & 6_1 & 0_1 & 1_1 & 2_1 & 3_1 & 4_2 & 5_2 & 6_2 & 0_2 & 1_2 & 2_2 & 3_2 \\ 0_2 & 1_2 & 2_2 & 3_2 & 4_2 & 5_2 & 6_2 & 0_1 & 1_1 & 2_1 & 3_1 & 4_1 & 5_1 & 6_1. \end{array} \tag{6.4}$$

The groups obtained by developing the initial group are given by the columns of the scheme

$$\begin{array}{ccccccc} 0_1 & 1_1 & 2_1 & 3_1 & 4_1 & 5_1 & 6_1 \\ 0_2 & 1_2 & 2_2 & 3_2 & 4_2 & 5_2 & 6_2. \end{array} \tag{6.5}$$

(c) The scope of the method of differences can be further extended by using the concept of "partial cycle" (P.C.), (cf. [14]). We shall illustrate the use of this concept by considering a specific example.

Let M be the module of residue classes mod (15), and to each element of M let there correspond a unique treatment. Consider the set of treatments (0,3,6,9,12). This set cannot form an initial group for the purposes of Theorem 2, since the differences arising are not all different but are the elements 3,6,9, 12 each repeated five times. We however note that if we develop this set, then the complete cycle of 15 sets consists of the three sets (0,3,6,9,12), (1,4,7,10,13) and (2,5,8,11,14) each repeated five times. We can therefore say that the complete cycle is divisible into 5 equal parts. If we take only a partial cycle, namely $\frac{1}{5}$ of

TABLE VI

GD designs which can be generated by the method of differences

Serial no.	Parameters v b r k / m n λ_1 λ_2	Initial group	Initial blocks	Modulus
(1)	14 28 6 3 7 2 0 1	$(0_1,0_2)$	$(1_1,6_1,0_2)$, $(2_1,5_1,0_2)$, $(3_1,4_1,0_2)$, $(1_2,2_2,4_2)$	mod (7)
(2)	14 14 4 4 7 2 0 1	$(0_1,0_2)$	$(1_1,2_1,4_1,0_2)$, $(1_2,2_2,4_2,0_1)$	mod (7)
(3)	26 52 8 4 13 2 0 1	$(0_1,0_2)$	$(1_1,3_1,9_1,0_2)$, $(2_1,6_1,5_1,0_2)$, $(1_2,3_2,9_2,0_1)$, $(2_2,6_2,5_2,0_1)$	mod (13)
(4)	18 54 9 3 9 2 2 1	$(0_1,0_2)$	$(0_1,3_1,1_2)$, $(0_1,4_2,0_2)$, $(0_1,5_2,8_2)$, $(0_1,6_2,7_2)$, $(0_1,1_1,4_1)$, $(0_1,2_1,2_2)$	mod (9)
(5)	30 75 10 4 15 2 2 1	$(0_1,0_2)$	$(0_1,2_1,14_1,4_2)$, $(0_2,2_2,14_2,4_1)$, $(0_1,4_1,10_1,1_2)$, $(0_2,4_2,10_2,1_1)$, $(0_1,8_1,0_2,8_2)$	mod (15)
(6)	39 78 10 5 13 3 2 1	$(0_1,0_2,0_3)$	$(1_1,3_1,9_1,0_2,0_3)$, $(2_1,6_1,5_1,0_2,0_3)$, $(1_2,3_2,9_2,0_3,0_1)$, $(2_1,6_2,5_2,0_3,0_1)$, $(1_3,3_3,9_3,0_1,0_2)$, $(2_3,6_3,5_3,0_1,0_2)$	mod (13)
(7)	10 20 8 4 5 2 0 3	$(0_1,0_2)$	$(0_1,1_2,2_2,4_2)$, $(0_2,1_1,2_1,4_1)$, $(0_1,2_2,3_2,4_2)$, $(0_2,2_1,3_1,4_1)$	mod (5)
(8)	16 32 6 3 4 4 0 1	(0,4,8,12) ¼ P.C.	(0,1,10), (0,2,5)	mod (16)
(9)	24 72 9 3 4 6 0 1	(0,4,8,12,16,20) ⅙ P.C.	(0,1,11), (0,2,7), (0,3,9)	mod (24)
(10)	15 30 6 3 5 3 0 1	(0,5,10) ⅓ P.C.	(0,6,8), (0,11,14)	mod (15)
(11)	15 45 9 3 3 5 2 1	(0,3,6,9,12) ⅕ P.C.	(0,6,12), (0,3,4), (0,2,7)	mod (15)
(12)	12 12 4 4 6 2 2 1	(0,6) ½ P.C.	(0,1,4,6)	mod (12)
(13)	12 36 9 3 4 3 0 2	(0,4,8) ⅓ P.C.	(0,1,3), (0,1,6), (0,2,5)	mod (12)

TABLE VI—*Cont.*

Serial no.	Parameters v b r k / m n λ_1 λ_2	Initial group	Initial blocks	Modulus
(14)	26 26 9 9 13 2 0 3	(0,13) ½ P.C.	(0,1,2,8,11,18,20,22,23)	mod (26)
(15)	35 70 10 5 5 7 2 1	(00,10,20,30,40,50,60) ⅐ P.C.	(10,20,40,01,04), (10,20,40,02,03)	mod (7, 5)
(16)	33 33 7 7 3 11 2 1	(00,10,20,30,40,50, 60,70,80,90,*t*0) 1/11 P.C.	(10,40,50,90,30,01,02)	mod (11, 3)
(17)	15 30 8 4 3 5 1 2	(00,10,20,30,40) ⅕ P.C.	(00,40,21,22), (00,20,11,12)	mod (5, 3)
(18)	15 30 10 5 5 3 2 3	(00,10,20) ⅓ P.C.	(00,10,21,22,24), (00,10,21,22,23)	mod (3, 5)
(19)	24 60 10 4 3 8 2 1	(00,30,60,90, 01,31,61,91) ⅛ P.C.	(00,10,40,91) C.C. (00,20,50,31) C.C. (00,60,01,61) ½ P.C.	mod (12, 2)
(20)	24 80 10 3 6 4 0 1	(00,20,40,60) ¼ P.C.	(00,10,61) C.C. (00,50,71) C.C. (00,11,42) C.C. (00,01,02) ⅓ P.C.	mod (8, 3)
(21)	12 30 10 4 3 4 2 3	(00,01,30,31) ¼ P.C.	(00,20,30,11) C.C. (00,10,50,41) C.C. (00,20,01,21) ½ P.C.	mod (6, 2)

the complete cycle for our groups, we see that any two treatments, the differences arising from which are 3,12 or 6,9, occur together just once in a group.

We now note that among the 18 differences arising from the initial blocks

(6.6) (0,6,12), (0,3,4), (0,2,7)

the elements 3,6,9,12 each occur twice, and the other nonzero elements, namely 1,2,4,5,7,8,10,11,13,14 each occur once. If, therefore, we develop these initial blocks mod (15) we get design (11) of Table VI, the groups consisting of ⅕ of the complete cycle obtained by developing the initial group (0,3,6,9,12). This is denoted by writing ⅕ P.C. after (0,3,6,9,12) in the column 3 of Table VI.

We may now state the following obvious generalization of Theorem 2.

THEOREM 3. *Let M be a module with cm elements and to each element of M let there correspond n/c treatments (c is supposed to be a divisor of n). Let it be possible to find t initial blocks each containing k treatments, and an initial group G containing n treatments such that:*

(i) *The differences arising from G consist of $n(n-1)/c$ different differences each repeated c times, the complete cycle of G being divisible into c equal parts.*

(ii) *Among the $k(k-1)t$ differences arising from the initial blocks each difference occurs λ_2 times, except the $n(n-1)/c$ differences arising from G, each of which occurs λ_1 times.*

Then by developing the initial blocks $B_1, B_2, \cdots, B_t$ we get the GD design with parameters $v = mn$, $b = mct$, $r = kct/n$, k, m, n, λ_1, λ_2, the groups being $1/c$th part of the complete cycle obtained by developing G.

In particular let $c = n$, and let M be the module of residue classes mod (mn), one treatment corresponding to each element of M. Let G be

$$(0, m, 2m, \cdots, m(n-1)). \tag{6.7}$$

Then the differences arising from G are the $n-1$ elements $m, 2m, \cdots, (n-1)m$ each repeated n times. The complete cycle of G is divisible into n equal parts and we can get $1/n$ part of this cycle, by adding $0, 1, \cdots, m-1$ to the elements of G and taking residues mod (mn). This gives us the m groups. If it is possible to find the initial blocks $B_1, B_2, \cdots, B_t$ each with k treatments, such that the differences arising from them consist of the elements $m, 2m, \cdots, (n-1)m$ each repeated λ_1 times, and all other nonzero elements of M each repeated λ_2 times, then by developing $B_1, B_2, \cdots, B_t$ we get the blocks of the GD design with parameters $v = mn$, $b = mnt$, $r = kt$, $m, n, \lambda_1, \lambda_2$. Designs (8)–(14) of Table VI have all been obtained in this manner.

(d) In applying the method of differences, the use of systems of double modulus (u, v) is often advantageous. The elements of such a system are binary symbols xy, where x is a residue class mod (u) and y is a residue class mod (v). In adding two elements, we add the components separately and reduce the first component mod (u) and the second component mod (v).

In applying Theorem 3, using systems of double modulus we shall take $u = n$, $v = m$, so that M is a system of double modulus (n, m). We shall illustrate by considering design (18) of Table VI, where $m = 5$, $n = 3$. The initial group G is (00,10,20), and consists of all elements of M for which the second component is zero. The complete cycle of G consists of 15 groups divisible into 3 equal parts. One of these parts is obtained by adding to G all the element of M for which the first component is zero. The groups of this "partial cycle" are taken as our groups They are given by the columns of

$$\begin{array}{ccccc} 00 & 01 & 02 & 03 & 04 \\ 10 & 11 & 12 & 13 & 14 \\ 20 & 21 & 22 & 23 & 24. \end{array} \tag{6.8}$$

The fact that the groups are obtained by taking only $\frac{1}{3}$ of the complete cycle obtainable from G is denoted by writing $\frac{1}{3}$ P.C. after (00,10,20) in column 3 of Table VI. The differences arising from G are all the nonnull elements of M for which the second component is zero, each repeated 3 times. If we now note

that among the forty differences arising from the initial blocks (00,10,21,22,24), (00,10,21,22,23) the elements 10,20 of M each occur twice, and the other nonnull elements of M each occur thrice, it follows from Theorem 3 that on developing these initial blocks we shall obtain all the blocks of design (18) of Table VI. Designs (15)–(18) of Table VI have all been obtained in this manner. In design (16), t stands for 10.

(e) Finally instead of considering only complete cycles developed from initial blocks, we may also allow partial cycles. This will be illustrated by considering design (20) of Table VI. M is here a system of double modulus (8,3). The initial group G consists of $n = 4$ elements (00,20,40,60). The differences arising from G are the elements 20,40,60 each occurring four times. For our groups we therefore take $\frac{1}{4}$ part of the complete cycle obtained by developing G. Our blocks should be such that two treatments differing by ± 20 or ± 40 should not be in the same block, but any two treatments the difference of which is anything else should occur in a block just once. Now the differences arising from the initial blocks (00,10,61), (00,50,71), (00,11,42) are all the elements of M (occurring once) except 20,40,60,01,02. Hence by developing these initial blocks we would get all pairs of treatments occurring together except those which differ by ± 20, ± 40, ± 01. We can therefore complete the solution by adding the initial block (00,01,02) and taking $\frac{1}{3}$ of the complete cycle obtainable from it, since the differences arising from it are 01 and 02 each repeated thrice. Designs (19) and (21) of Table V have also been obtained in a similar manner. The letters C.C. after an initial block mean that we have to take the complete cycle developed from it, whereas $1/n$ P.C. after an initial block means that only $1/n$ part of the complete cycle has to be taken. Of course this notation has been used only for those designs in which some of the initial blocks have partial cycles.

It should be noted that Theorem 3 when properly interpreted remains valid even when some of the initial blocks have partial cycles. If $1/s$ part of the cycle arising from a block is taken, then this block counts only as $1/s$ blocks, and the differences arising from it count only as $k(k-1)/s$ differences (i.e., every set of s identical differences counts only as one). Thus in design (20) of Table VI, the number of initial blocks is $t = 10\frac{1}{3}$ since only $\frac{1}{3}$ of the cycle of the last initial block is taken. Since to each element there corresponds only one treatment $c = n$, the relation $r = kct/n$ is seen to remain valid. The $k(k-1)t$ differences arising from the initial blocks are the 6×3 differences arising from the first three initial blocks, together with the two differences arising from the last initial block.

7. Construction of semi-regular GD designs with $\lambda_1 = 0$.

(a) For a semi-regular GD design $P = rk - v\lambda_2 = 0$ by definition. Hence from (2.0) and (2.1)

$$r = \lambda_2 n - \lambda_1(n-1). \tag{7.0}$$

In this section we shall consider the case $\lambda_1 = 0$. This leads to $r = \lambda_2 n$, $k = m$. Hence the parameters of the design can be written as

$$v = mn, \quad b = n^2\lambda_2, \quad r = n\lambda_2, \quad k = m, \quad m, n, \lambda_1 = 0, \lambda_2. \tag{7.1}$$

We shall first establish the equivalence of the design (7.1) with an orthogonal array $A = [\lambda_2 n^2, m, n, 2]$ of strength 2, which may be defined as a matrix $A = (a_{ij})$, with m rows and $\lambda_2 n^2$ columns for which each element a_{ij} is one of the integers $0, 1, 2, \cdots, n-1$, and which has the orthogonality property that for any two rows, say i and u, the pairs $(a_{ij}, a_{uj}), j = 1, 2, \cdots, \lambda_2 n^2$ occurring in the corresponding columns consist of all possible ordered pairs of the integers 0, 1, $2, \cdots, n-1$, each repeated λ_2 times. It follows that each of the integers $0, 1, 2, \cdots, n-1$ appears $n\lambda_2$ times in each row of A. Orthogonal arrays have been studied by Plackett and Burman, Rao, Bush and one of the authors (Bose), [16], [17], [18], [19], [20], [21], [22].

THEOREM 4. *The existence of a semi-regular GD design with parameters* (7.1) *implies the existence of an orthogonal array* $A = [\lambda_2 n^2, m, n, 2]$ *of strength* 2, *and conversely.*

PROOF. Replace any integer x appearing in the ith row of A by the treatment $(i-1)n + x$. The ith row of the derived scheme now contains the treatments

$$(i-1)n, (i-1)n+1, \cdots, (i-1)n+n-1. \tag{7.15}$$

We shall show that the columns of the derived scheme give the blocks of the GD design (7.1), where the ith group of treatments is (7.15) Treatments belonging to the ith group occur only in the ith row of the derived scheme. Hence two treatments belonging to different groups never occur together in the same block (column). Also from the orthogonality property of A it follows that any two treatments belonging to different groups occur together in λ_2 blocks. This proves our statement.

Conversely, suppose there exists a semi-regular GD design with parameters (7.1). Let the ith group of treatments be given by (7.15), $i = 1, 2, \cdots, m$. It has been shown in [2] that each block of a semi-regular GD design contains the same number of treatments from each group. Since $k = m$ in the present case, each block contains just one treatment from each block. We can now exhibit the blocks of (7.1) as the columns of a rectangular scheme in which the treatments of the ith group occupy the ith row. Replacing the treatment

$$(i-1)n + x$$

of the ith group by x, $x = 1, 2, \cdots, n-1$, $i = 1, 2, \cdots, m$. We then get an orthogonal array A of size $\lambda_2 n^2$, m constraints, n levels and strength 2. This proves the equivalence of the orthogonal array A and the GD design (7.1).

COROLLARY. *The existence of GD design* (7.1) *implies the existence of the GD design with parameters*

$$\begin{aligned} &v = m_1 n, \quad b = n^2\lambda_2, \quad r = n\lambda_2, \quad k = m_1, \\ &m_1, \quad n, \quad \lambda_1 = 0, \quad \lambda_2 = 1 \end{aligned} \tag{7.2}$$

where $m_1 < m$.

If the GD design (7.1) is written in a form in which the columns give the blocks, and the treatments of the ith group appear only in the ith row, then to get the blocks of (7.2), we have simply to discard the last $m - m_1$ rows.

(b) In special cases the blocks of GD designs with parameters (7.1) can be obtained more expeditiously by using affine resolvable BIB designs or finite geometries rather than by directly using orthogonal arrays.

A resolvable BIB design is said to be affine resolvable if any two blocks of different replications have exactly the same number of treatments in common. It has been shown by one of the authors (Bose) [10], that the necessary and sufficient condition for a resolvable BIB design to be affine resolvable is

$$b^* = v^* + r^* - 1. \tag{7.25}$$

In this case the number of treatments common to blocks of two different replications is k^{*2}/v^*, which must therefore be integral. The connection between orthogonal arrays and affine resolvable BIB designs was noticed by Plackett and Burman [19].

It is clear that if we dualize an affine resolvable BIB design with parameters $v^*, b^*, r^*, k^*, \lambda^*$, we get a semi-regular GD design with parameters

$$\begin{gathered} v = b^*, \quad b = v^*, \quad r = k^*, \quad k = r^*, \\ m = r^*, \quad n = b^*/r^*, \quad \lambda_1 = 0, \quad \lambda_2 = k^{*2}/v^*. \end{gathered} \tag{7.3}$$

In particular the BIB designs (3.3) belonging to the series $0S\ 1$ are affine resolvable and lead by dualization to the blocks of the GD design

$$\begin{gathered} v = s^2 + s, \quad b = s^2, \quad r = s, \quad k = s + 1, \\ m = s + 1, \quad n = s, \quad \lambda_1 = 0, \quad \lambda_2 = 1. \end{gathered} \tag{7.35}$$

From this we can get the blocks for (cf. Theorem 4, Corollary)

$$\begin{gathered} v = ms, \quad b = s^2, \quad r = s, \quad k = m, \\ m, \quad n = s, \quad \lambda_1 = 0, \quad \lambda_2 = 1, \end{gathered} \tag{7.4}$$

where $m < s + 1$.

It will appear that we can express the blocks of (7.4) in a resolvable form. This will be illustrated by considering the special case $s = 4$. The columns of scheme (3.5) give the blocks of the BIB design $v^* = 16, b^* = 20, r^* = 5, k^* = 4, \lambda^* = 1$ in a resolvable form. Let us write down the dual of this design. The blocks of the dual corresponding to the treatments of the original can now be numbered 0, 1, 2, $\cdots$, 14 and ∞. Also the treatments of the dual corresponding to the blocks of the original can be numbered 1, 2, $\cdots$, 20, and can be divided into five groups corresponding to the replications. If in the original (3.5), the treatment i occurs in the block j in the dual we put the treatment j in the block i. The blocks of the dual are then given by the columns of the following scheme, where the last column corresponds to the block ∞.

$$\begin{array}{cccccccccccccccc} 4 & 1 & 2 & 1 & 1 & 4 & 2 & 3 & 2 & 2 & 4 & 3 & 1 & 3 & 3 & 4 \\ 7 & 8 & 5 & 6 & 5 & 5 & 8 & 6 & 7 & 6 & 6 & 8 & 7 & 5 & 7 & 8 \\ 11 & 11 & 12 & 9 & 10 & 9 & 9 & 12 & 10 & 11 & 10 & 10 & 12 & 11 & 9 & 12 \\ 13 & 15 & 15 & 16 & 13 & 14 & 13 & 13 & 16 & 14 & 15 & 14 & 14 & 16 & 15 & 16 \\ 19 & 17 & 19 & 19 & 20 & 17 & 18 & 17 & 17 & 20 & 18 & 19 & 18 & 18 & 20 & 20. \end{array} \tag{7.45}$$

Finally, we rearrange the blocks so that all blocks containing the same treatment of the last group come together, and arrive at the scheme

(7.5)

1	4	3	2	2	4	1	3	4	2	1	3	1	2	3	4
8	5	6	7	8	6	7	5	7	5	6	8	5	6	7	8
11	9	12	10	9	10	12	11	11	12	9	10	10	11	9	12
15	14	13	16	13	15	14	16	13	15	16	14	13	14	15	16
17	17	17	17	18	18	18	18	19	19	19	19	20	20	20	20.

Taking only the first m rows of the scheme (7.5) the columns give the blocks of the semi-regular GD design

(7.55)
$$v = 4m, \quad b = 16, \quad r = 4, \quad k = m,$$
$$m, \quad n = 4, \quad \lambda_1 = 0, \quad \lambda_2 = 1$$

when $m < 5$, the design is in a resolvable form the replications being separated by the vertical lines.

(c) The connection between orthogonal arrays and finite geometries is given in [22]. We shall now illustrate the use of finite geometries in obtaining the blocks of semi-regular GD designs.

Consider the finite projective geometry PG(3, p^n), where p is a prime, and set $s = p^n$. There are exactly $s^2 + s + 1$ lines passing through any point O. Let us choose O = (0,0,0,1). Choose any $m \leqq s^2 + s + 1$ lines through O, and let the points other than O on these lines correspond to the treatments. We then have ms treatments divided into m groups, the s treatments corresponding to points on the same line forming a group. There are s^3 planes not passing through O. Each of these planes intersects a line through O in a unique point. Hence if we take these planes for blocks, then each block would contain exactly one treatment from each group. Also any treatment is contained in s^2 blocks. Two treatments belonging to the same group do not occur together in any block, but the points corresponding to two treatments of different groups are joined by a line through which s of the planes chosen for blocks pass. Hence two treatments not belonging to the same group occur together in s blocks. We thus get a semi-regular GD design with parameters

(7.6)
$$v = ms, \quad b = s^3, \quad r = s^2, \quad k = m,$$
$$m, \quad n = s, \quad \lambda_1 = 0, \quad \lambda_2 = s;$$

where $m \leqq s^2 + s + 1$.

We shall now show that if $m \leqq s^2$, then the blocks can be obtained in a resolvable form. Choose any plane through O, say $x_3 = 0$, and call it the fundamental plane. There are s^2 lines on the fundamental plane not passing through O. Through each of these lines there pass s planes chosen as blocks, which obviously give a complete replication provided that none of the m lines, the points of which (other than O) give the treatments, lie on the fundamental plane. Since there are s^2 lines through O not lying on the fundamental plane, we can get the blocks of (7.6) in a resolvable form if $m \leqq s^2$.

Again if $s^2 < m \leqq s^2 + s$, we can divide the blocks into s sets of s^2 each, such that the blocks of any set give s complete replications. This can be done by taking a fundamental line, say $x_2 = 0$, $x_3 = 0$. Let the lines whose points correspond to the treatments be different from the fundamental line. Then the s^2 blocks corresponding to planes passing through the same point of the fundamental line give s complete replications.

The equation of any plane not passing through O may be put in the form $ax_1 + bx_2 + cx_3 + x_4 = 0$ where a, b, c are elements of the Galois field $GF(p^n)$. Varying a, b, c we get all the s^3 planes. The s planes, for which a and b remain fixed but c takes the s different possible values, give a complete replication (when none of the lines, whose points correspond to the treatments, lie in $x_3 = 0$), and the s^2 planes, for which a remains fixed, but b and c take all possible values, give a set of s complete replications (when $x_2 = 0, x_3 = 0$ is not one of the lines whose points correspond to the treatments). After the blocks have been calculated the points representing the treatments may be identified with the treatments $1, 2, \cdots, ms$.

Using PG(3, 2) we find that, if we retain only the first m rows of the scheme (7.7), then the columns represent the 8 blocks of the semi-regular GD design

$$\begin{aligned} &v = 2m, \quad b = 8, \quad r = 4, \quad k = m, \\ &\quad m, \quad n = 2, \quad \lambda_1 = 0, \quad \lambda_2 = 2. \end{aligned} \tag{7.65}$$

The vertical lines separate the replications.

(7.7)

1	2	1	2	1	2	1	2
3	4	3	4	4	3	4	3
5	6	6	5	5	6	6	5
7	8	8	7	8	7	7	8
9	9	10	10	9	9	10	10
11	11	12	12	12	12	11	11
13	13	13	13	14	14	14	14

The groups for (7.65) are given by the first m columns of

$$\begin{array}{cccccccc} 1 & 3 & 5 & 7 & 9 & 11 & 13 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14. \end{array} \tag{7.75}$$

Similarly using PG(3, 3) we find that, if we retain only the first m rows of the scheme (7.83), then the columns represent the 27 blocks of the semi-regular GD design

$$\begin{aligned} &v = 3m, \quad b = 27, \quad r = 9, \quad k = m, \\ &\quad m, \quad n = 3, \quad \lambda_1 = 0, \quad \lambda_2 = 3. \end{aligned} \tag{7.8}$$

As before the vertical lines separate the replications.

(7.83)

1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
4	5	6	4	5	6	4	5	6	5	6	4	5	6	4	5	6	4	6	4	5	6	4	5	6	4	5
7	8	9	7	8	9	7	8	9	9	7	8	9	7	8	9	7	8	8	9	7	8	9	7	8	9	7
10	11	12	11	12	10	12	10	11	10	11	12	11	12	10	12	10	11	10	11	12	11	12	10	12	10	11
13	14	15	14	15	13	15	13	14	14	15	13	15	13	14	13	14	15	15	13	14	13	14	15	14	15	13
16	17	18	17	18	16	18	16	17	18	16	17	16	17	18	17	18	16	17	18	16	18	16	17	16	17	18
19	20	21	21	19	20	20	21	19	19	20	21	21	19	20	20	21	19	19	20	21	21	19	20	20	21	19
22	23	24	24	22	23	23	24	22	23	24	22	22	23	24	24	22	23	24	22	23	23	24	22	22	23	24
25	26	27	27	25	26	26	27	25	27	25	26	26	27	25	25	26	27	26	27	25	25	26	27	27	25	26
28	28	28	29	29	29	30	30	30	28	28	28	29	29	29	30	30	30	28	28	28	29	29	29	30	30	30
31	31	31	32	32	32	33	33	33	32	32	32	33	33	33	31	31	31	33	33	33	31	31	31	32	32	32
34	34	34	36	36	36	35	35	35	35	35	35	34	34	34	36	36	36	36	36	36	35	35	35	34	34	34
37	37	37	37	37	37	37	37	37	38	38	38	38	38	38	38	38	38	39	39	39	39	39	39	39	39	39

The groups for (7.8) are given by the first m columns of

(7.86)

1	4	7	10	13	16	19	22	25	28	31	34	37
2	5	8	11	14	17	20	23	26	29	32	35	38
3	6	9	12	15	18	21	24	27	30	33	36	39.

(d) Whenever an orthogonal array $[\lambda_2 n^2, m, n, 2]$ of strength 2 is directly available we can use it for obtaining the blocks of (7.1). The procedure to be followed has already been explained in the proof of Theorem 4.

Using the array [18, 7, 3, 2] given in [22], we get the blocks of the semi-regular GD design

$$v = 3m, \quad b = 18, \quad r = 6, \quad k = m, \tag{7.9}$$
$$m, \quad n = 3, \quad \lambda_1 = 0, \quad \lambda_2 = 2;$$

where $m \leqq 7$, by retaining only the first m rows of the scheme.

(7.92)

1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
4	5	6	4	5	6	5	6	4	6	4	5	5	6	4	6	4	5
7	8	9	8	9	7	7	8	9	9	7	8	9	7	8	8	9	7
10	11	12	12	10	11	12	10	11	10	11	12	11	12	10	11	12	10
13	14	15	14	15	13	15	13	14	14	15	13	13	14	15	15	13	14
16	17	18	18	16	17	17	18	16	17	18	16	18	16	17	16	17	18
19	19	19	19	19	19	20	20	20	20	20	20	21	21	21	21	21	21.

As before, the blocks are given by columns, and the vertical lines divide complete replications. Thus the design is resolvable for $m \leqq 6$. The groups are given by the first m columns of

(7.94)

1	4	7	10	13	16	19
2	5	8	11	14	17	20
3	6	9	12	15	18	21.

Similarly using the array [32,9,4,2] given in [22] we can get the blocks of the semi-regular GD design

$$(7.96)\qquad \begin{aligned} v &= 4m, \quad & b &= 32, \quad & r &= 8, \quad & k &= m, \\ m, \quad & & n &= 4, \quad & \lambda_1 &= 0, \quad & \lambda_2 &= 2 \end{aligned}$$

if $m \leqq 9$. The design can be obtained in a resolvable form if $m \leqq 8$.

Plackett and Burman [19] have given orthogonal arrays $[4\lambda, 4\lambda - 1, 2, 2]$ for all integral $\lambda \leqq 25$, except $\lambda = 23$. These may be used to obtain the blocks of the corresponding singular GD designs with parameters

TABLE VII

Parameters of semi-regular GD designs with $\lambda_1 = 0, \lambda_2 \leqq 3, r \leqq 10$

Serial no.	Parameters								Maximum m	Maximum m for resolvability
	v	b	r	k	m	n	λ_1	λ_2		
(1)	$3m$	9	3	m	m	3	0	1	4	3
(2)	$4m$	16	4	m	m	4	0	1	5	4
(3)	$5m$	25	5	m	m	5	0	1	6	5
(4)	$6m$	36	6	m	m	6	0	1	3	2
(5)	$7m$	49	7	m	m	7	0	1	8	7
(6)	$8m$	64	8	m	m	8	0	1	9	8
(7)	$9m$	81	9	m	m	9	0	1	10	9
(8)	$10m$	100	10	m	m	10	0	1	3	2
(9)	$2m$	8	4	m	m	2	0	2	7	4
(10)	$3m$	18	6	m	m	3	0	2	7	6
(11)	$4m$	32	8	m	m	4	0	2	9	8
(12)	$5m$	50	10	m	m	5	0	2	6	5
(13)	$2m$	12	6	m	m	2	0	3	11	2
(14)	$3m$	27	9	m	m	3	0	3	13	9

$$(7.98)\qquad \begin{aligned} v &= 2m, \quad & b &= 4\lambda, \quad & r &= 2\lambda, \quad & k &= m, \\ m, \quad & & n &= 2, \quad & \lambda_1 &= 0, \quad & \lambda_2 &= \lambda; \end{aligned}$$

where $m \leqq 4\lambda - 1$. Of course only small values of λ and m yield designs of practical interest.

We present in Table VII the parameters of semi-regular GD designs for which $r \leqq 10, \lambda_1 = 0, \lambda_2 \leqq 3$, and the blocks for which can be obtained by the methods discussed in this section. The parameter m has been kept arbitrary, but the maximum value of m for which the design exists and also the maximum value of m for which the design can be obtained in a resolvable form has been given.

Number (12) is the duplicate of number (3), that is, is obtained by repeating each block of (3) twice. Numbers (4) and (8) can be obtained by first writing down the orthogonal array $[n^2, 3, n, 2]$ corresponding to an $n \times m$ Latin square

$n = 6, 10$, as it is well known that a set of $m - 2$ mutually orthogonal $n \times n$ Latin squares is equivalent to an orthogonal array $[n^2, m, n, 2]$, (cf. [18], [21]).

8. Construction of semi-regular GD designs for which $\lambda_1 \neq 0$, $\lambda_2 \neq 0$. Now $l' = rk - v\lambda_2 = 0$ by definition, and $k = cm$ since each block contains the same number of treatments from each group [2]. Using (2.0) and (2.1), the eight parameters of the design can be expressed in terms of m, n, λ_2 and c only. Thus the parameters are

$$v = mn, \quad b = n^2\lambda_2/c^2, \quad r = n\lambda_2/c, \quad k = cm, \tag{8.0}$$

$$m, n, \lambda_1 = n(c-1)\lambda_2/(n-1)c, \lambda_2. \tag{8.1}$$

Also as proved in [2] for a semi-regular GD design,

$$b \geqq v - m + 1. \tag{8.2}$$

$$m \leqq \frac{b-1}{n-1} = \frac{n^2\lambda_2 - c^2}{c^2(n-1)}. \tag{8.3}$$

TABLE VIIIA

Parameters of semi-regular GD designs with $\lambda_1 \neq 0$, $r \leqq 10$

Serial no.	Parameters								Maximum m
	v	b	r	k	m	n	λ_1	λ_2	
(1)	$4m$	12	6	$2m$	m	4	2	3	3
(2)	$3m$	9	6	$2m$	m	3	3	4	4
(3)	$6m$	20	10	$3m$	m	6	4	5	3

The values of n, c and λ_2 must be such as to make b, r and λ_1 integral, but m may be any integer subject to (8.3). It follows that, if $\lambda_1 \neq 0$, the only semi-regular GD designs in the range $r \leqq 10$ are those listed in Table VIIIA.

It is clear that, if the blocks and groups for the above designs can be obtained for the maximum value of m, then for any smaller value of m we have only to discard some of the groups and the treatments belonging to them. The groups and blocks for the designs in Table VIIIA are given in Table VIIIB (for the maximum value of m).

Here the groups have been given in full, and only the blocks have to be developed. The validity of the solution follows from the notion of differences developed in [8] and explained in section 6(a) of the present paper. For illustration we shall consider design (3) of Tables VIIIA and B, when m has the maximum value 3, and prove that the initial blocks shown give rise to it when developed.

The 18 treatments form three groups shown in the 2nd column of Table VIIIB. The 15 treatments other than ∞_1, ∞_2, ∞_3 fall into three classes according to the suffix carried (cf. section 6(a)). We shall distinguish three different types of pairs.

(i) Pairs of the type (∞_i, ∞_j); $i \neq j$; $i, j = 1, 2, 3$. Each of the three pairs (∞_1, ∞_2), (∞_2, ∞_3), (∞_3, ∞_1) occurs in just one initial block shown in the 3rd column of Table VIIIB. Since ∞ and the suffixes remain invariant when the blocks are developed, each of these pairs occurs five times in the completed design, as it should since $\lambda_2 = 5$ and ∞_i and ∞_j $(i \neq j)$ belong to different groups.

(ii) Pairs of the type (∞_i, u_j); $i, j = 1, 2, 3$; where u is an element of the field of residue classes, mod (5). When developed, the pair (∞_i, u_j) gives rise to five pairs, of which one component is ∞_i and the second component varies over all the five treatments of the jth class. In the initial blocks, ∞_i occurs with just 4 treatments of the jth class, if $i \neq j$, and 5 elements of the jth class, if $i \neq j$. It follows that any pair (∞_i, u_j) occurs 4 times in the completed design if $i = j$

TABLE VIIIB

Blocks and groups for semi-regular GD designs with $\lambda_1 \neq 0$, $r \leqq 10$

Serial no.	Groups	Initial blocks	Modulus
(1)	$(00_1,01_1,10_1,11_1)$ $(00_2,01_2,10_2,11_2)$ $(00_3,01_3,10_3,11_3)$	$(00_1,01_1;\ 00_2,10_2;\ 00_3,11_3)$ $(00_1,11_1;\ 00_2,01_2;\ 00_2,10_3)$ $(00_1,10_1;\ 00_2,11_2;\ 00_3,01_3)$	mod (2, 2)
(2)	$(0_1,1_1,2_1)$ $(0_2,1_2,2_2)$ $(0_3,1_3,2_3)$ $(\infty_1,\infty_2,\infty_3)$	$(0_1,1_1;\ 0_2,2_2;\ 0_3,2_3;\ \infty_2,\infty_3)$ $(0_1,2_1;\ 0_2,1_2;\ 0_3,2_3;\ \infty_3,\infty_1)$ $(0_1,2_1;\ 0_2,2_2;\ 0_3,1_3;\ \infty_1,\infty_2)$	mod (3)
(3)	$(0_1,1_1,2_1,3_1,4_1,\infty_1)$ $(0_2,1_2,2_2,3_2,4_2,\infty_2)$ $(0_3,1_3,2_3,3_3,4_3,\infty_3)$	$(0_1,1_1,2_1;\ 1_2,3_2,4_2;\ 0_3,1_3,2_3)$ $(\infty_1,3_1,4_1;\ \infty_2,0_2,2_2;\ 0_3,1_3,2_3)$ $(\infty_1,0_1,2_1;\ 0_2,1_2,2_2;\ \infty_3,0_3,2_3)$ $(1_1,3_1,4_1;\ \infty_2,3_2,4_2;\ \infty_3,0_3,2_3)$	mod (5)

and 5 times if $i \neq j$. This is as it should be, since ∞_i and u_j do or do not belong to the same group according as $i = j$ or $i \neq j$ and $\lambda_1 = 4$, $\lambda_2 = 5$.

(iii) Pairs of the type of (u_i, u_j); $i, j = 1, 2, 3$, where u is an element of the field of residue classes mod (5). It can be verified that leaving out ∞_1, ∞_2, ∞_3 the initial blocks give rise to each pure difference 4 times and each mixed difference 5 times. Hence in the completed design any pair (u_i, u_j) occurs 4 times if $i = j$ and 5 times if $i \neq j$, as it should, since u_i and u_j do or do not belong to the same group according as $i = j$ or $i \neq j$.

Again it is easy to see that each of the treatments ∞_1, ∞_2, ∞_3 occurs 10 times in the completed design, since each of these occurs twice in the initial blocks. The other treatments also occur 10 times in the completed design, since each class is represented 10 times in the initial blocks. This completes the proof.

If, in design (3) of Table VIIIA, $m = 2$, then the corresponding blocks can be obtained by developing the initial blocks shown in Table VIIIB, after drop-

ping the treatments with suffix 3. It should be noted that the first two initial blocks now give a complete replication, and the same is true of the last two initial blocks. Hence the blocks are obtained in a resolvable form.

9. GD designs derivable by replication addition and subtraction. Consider a BIB design with parameters v^*, b^*, r^*, k^*, λ^* in which v^* is divisible by k^*, and suppose that either a resolvable solution is known, or at least a solution is known in a form where there are v^*/k^* blocks which give a complete replication. Then we can get a GD design with parameters

$$(9.0) \quad v = v^*, \quad b = tb^* + a(v^*/k^*), \quad r = tr^* + a, \quad k = k^*$$

$$(9.1) \quad m = v^*/k^*, \quad n = k^*, \quad \lambda_1 = t\lambda^* + a, \quad \lambda_2 = t\lambda^*$$

in the following manner. Choose a set of v^*/k^* blocks giving a complete replication. Repeat the BIB design t times, and then add the chosen set of blocks a times. Then we get a GD design with parameters given by (9.0) and (9.1), for which the groups are given by the chosen set of blocks.

When the BIB design is repeated t times, the chosen set of blocks is also repeated t times. Hence instead of adding the chosen set of blocks a times, we could delete the chosen set of blocks a_1 times ($a_1 \leqq t$). This would give a GD design with parameters (9.0) and (9.1) with $a = -a_1$. If the original BIB design is resolvable, then the derived GD design is also resolvable.

For example, if we start with the BIB designs of the series $0S$ 1 whose parameeters are given by (3.3), we get resolvable GD designs with parameters

$$(9.2) \quad v = s^2, \quad b = t(s^2 + s) + as, \quad r = t(s + 1) + a, \quad k = s$$

$$(9.3) \quad m = s, \quad n = s \quad \lambda_1 = t + a, \quad \lambda_2 = t$$

where $a \geqq -t$, and s is a prime or a prime power. As an illustration let $s = 4$, $t = 1, a = -1$. The blocks of the BIB design $v^* = 16$, $b^* = 20$, $r^* = 5$, $k^* = 4$, $\lambda^* = 1$ are given in a resolvable form by (3.5). Hence the blocks of the GD design with parameters

$$v = 16, \quad b = 16, \quad r = 4, \quad k = 4,$$

$$m = 4, \quad n = 4, \quad \lambda_1 = 0, \quad \lambda_2 = 1$$

are obtained by taking any four replications from (3.5); the remaining replication then gives the groups.

The blocks of BIB designs belonging to the series $0S$ 1 can be obtained in a resolvable form as explained in Section 3, by using the difference sets in Table I. The blocks for all other BIB designs occurring in Table IX can be found in Table III, being in a resolvable form in every case except $v^* = 45$, $b^* = 99$, $r^* = 11$, $k^* = 5$, $\lambda^* = 1$. In this case the block $(00_1, 00_2, 00_3, 00_4, 00_5)$, when developed mod (3, 3), provides a complete replication.

10. Extension of GD designs. Suppose that there exists a resolvable group divisible design with parameters

$$v = k\alpha, \quad b = r\alpha, \quad r, \ k, \ m, \ n, \ \lambda_1, \ \lambda_2 = 1 \tag{10.0}$$

TABLE IX

Parameters of GD designs derivable from BIB designs by replication addition or subtraction

Serial no. and series	Parameters of BIB design					Auxiliary parameters		Parameters of GD design							
	v^*	b^*	r^*	k^*	λ^*	t	a	v	b	r	k	m	n	λ_1	λ_2
(1) $0S\ 1+$	16	20	5	4	1	1	−1	16	16	4	4	4	4	0	1
(2) $0S\ 1+$	16	20	5	4	1	1	1	16	24	6	4	4	4	2	1
(3) $0S\ 1+$	16	20	5	4	1	1	2	16	28	7	4	4	4	3	1
(4) $0S\ 1+$	16	20	5	4	1	2	−2	16	32	8	4	4	4	0	2
(5) $0S\ 1+$	16	20	5	4	1	2	−1	16	36	9	4	4	4	1	2
(6) $0S\ 1+$	25	30	6	5	1	1	−1	25	25	5	5	5	5	0	1
(7) $0S\ 1+$	25	30	6	5	1	1	1	25	35	7	5	5	5	2	1
(8) $0S\ 1+$	25	30	6	5	1	1	2	25	40	8	5	5	5	3	1
(9) $0S\ 1+$	25	30	6	5	1	2	−2	25	50	10	5	5	5	0	2
(10) $0S\ 1+$	49	56	8	7	1	1	−1	49	49	7	7	7	7	0	1
(11) $0S\ 1+$	49	56	8	7	1	1	1	49	63	9	7	7	7	2	1
(12) $0S\ 1+$	49	56	8	7	1	1	2	49	70	10	7	7	7	3	1
(13) $0S\ 1+$	64	72	9	8	1	1	−1	64	64	8	8	8	8	0	1
(14) $0S\ 1+$	64	72	9	8	1	1	1	64	72	10	8	8	8	2	1
(15) $0S\ 1+$	81	90	10	9	1	1	−1	81	81	9	9	9	9	0	1
(16) T_1+	15	35	7	3	1	1	−1	15	30	6	3	5	3	0	1
(17) T_1+	15	35	7	3	1	1	1	15	40	8	3	5	3	2	1
(18) T_1+	15	35	7	3	1	1	2	15	45	9	3	5	3	3	1
(19) F_2+	28	63	9	4	1	1	−1	28	54	8	4	7	4	0	1
(20) F_2+	28	63	9	4	1	1	1	28	69	10	4	7	4	2	1
(21) T_1+	21	70	10	3	1	1	−1	21	63	9	3	7	3	0	1
(22) G_2	45	99	11	5	1	1	−1	45	90	10	5	9	5	0	1

so that the b blocks are divisible into r sets of α blocks, each set giving a complete replication. Let

$$v' = r, \quad b' = \frac{r(r-\alpha)}{k+1}, \quad r' = r - \alpha, \quad k' = k + 1,$$
$$m' = \frac{r}{n}, \quad n' = n, \quad \lambda_1' = \lambda_1, \quad \lambda_2' = 1. \tag{10.1}$$

Then clearly

$$v' = m'n', \qquad b'k' = v'r', \tag{10.2}$$

and it follows from (2.0) and (2.1) that

$$\lambda_1'(n' - 1) + \lambda_2'n'(m' - 1) = r'(k' - 1). \tag{10.25}$$

Hence if b' and m' are integers, the parameters $v', b', r', k', m', n', \lambda_1', \lambda_2'$ given by (10.1) can be the parameters of a GD design. Suppose a combinatorial solution of this design is available. We shall show that in this case we can build up a solution of the GD design with parameters

$$\begin{aligned} &v'' = v + v', \quad b'' = b + b', \quad r'' = r, \quad k'' = k + 1, \\ &m'' = m + m', \quad n'' = n, \quad \lambda_1'' = \lambda_1, \quad \lambda_2'' = 1. \end{aligned} \tag{10.3}$$

Let the treatments in (10.0) and (10.1) be different so that there are altogether $v + v'$ treatments. To each block in the ith replication of (10.0) adjoin the ith treatment of (10.1), $(i = 1, 2, \cdots, r)$. To the design (10.0) so extended, add all the blocks of (10.1). This gives us a combinatorial solution of (10.3) where the groups are the groups of (10.0) and (10.1) taken together. It is easy to see that the necessary conditions are satisfied. This method may be called the method of extension.

As an illustration we shall build up the solution of the GD design

$$\begin{aligned} &v = 12, \quad b = 24, \quad r = 6, \quad k = 3, \\ &m = 6, \quad n = 2, \quad \lambda_1 = 2, \quad \lambda_2 = 1, \end{aligned} \tag{10.4}$$

starting from a solution of

$$\begin{aligned} &v = 6, \quad b = 18, \quad r = 6, \quad k = 2, \\ &m = 3, \quad n = 2, \quad \lambda_1 = 2, \quad \lambda_2 = 1 \end{aligned} \tag{10.45}$$

which can be obtained by adding one complete replication, say the last, to the solution (3.1) of the BIB design (3.0). Here $\alpha = 3$, and we see from (10.1) that for extension we require a solution of

$$\begin{aligned} &v' = 6, \quad b' = 6, \quad r' = 3, \quad k' = 3, \\ &m' = 3, \quad n' = 2, \quad \lambda_1' = 2, \quad \lambda_2' = 1. \end{aligned} \tag{10.5}$$

It is seen from Theorem 2 that a solution of this is obtainable by developing mod (6) the initial block (0,1,3). However to keep the treatments of (10.5) distinct from those of (10.45) we may replace the ith treatment of (10.5) by a_i. Proceeding as explained, the blocks of (10.4) are given by the columns of the scheme

$$\begin{array}{cccccccccccccccccccccccc} 1 & 2 & 0 & 2 & 3 & 1 & 3 & 4 & 2 & 4 & 0 & 3 & 0 & 1 & 4 & 0 & 1 & 4 & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \\ 4 & 3 & \infty & 0 & 4 & \infty & 1 & 0 & \infty & 2 & 1 & \infty & 3 & 2 & \infty & 3 & 2 & \infty & a_1 & a_2 & a_3 & a_4 & a_5 & a_0 \\ a_0 & a_0 & a_0 & a_1 & a_1 & a_1 & a_2 & a_2 & a_2 & a_3 & a_3 & a_3 & a_4 & a_4 & a_4 & a_5 & a_5 & a_5 & a_3 & a_4 & a_5 & a_0 & a_1 & a_2, \end{array} \tag{10.6}$$

and the groups are given by the columns of

$$\begin{matrix} 0 & 1 & 4 & a_0 & a_1 & a_2 \\ 3 & 2 & \infty & a_3 & a_4 & a_5 \end{matrix} \tag{10.65}$$

Again we can build up the solution of the GD design

$$\begin{aligned} v &= 24, & b &= 54, & r &= 9, & k &= 4, \\ m &= 8, & n &= 3, & \lambda_1 &= 3, & \lambda_2 &= 1 \end{aligned} \tag{10.7}$$

by starting with the design

$$\begin{aligned} v &= 15, & b &= 45, & r &= 9, & k &= 3, \\ m &= 5, & n &= 3, & \lambda_1 &= 3, & \lambda_2 &= 1 \end{aligned} \tag{10.8}$$

which is design (18) of Table IX, and use for extension the solution of

$$\begin{aligned} v' &= 9, & b' &= 9, & r' &= 4, & k' &= 4, \\ m' &= 3, & n' &= 3, & \lambda_1' &= 3, & \lambda_2' &= 1 \end{aligned} \tag{10.9}$$

which can be obtained by developing mod (9) the initial block (0,1,3,6).

11. Addition of GD designs. The method of addition consists of getting a new GD design by taking together the blocks of two suitable GD designs with the same v and k. It may be regarded as a slight generalization of the method of replication addition discussed in Section 9. This will be explained by two examples.

(a) If in (7.4) we put $m = s - 1$, we get the GD design

$$\begin{aligned} v &= s^2 - s, & b &= s^2, & r &= s, & k &= s - 1, \\ m &= s - 1, & n &= s, & \lambda_1 &= 0, & \lambda_2 &= 1, \end{aligned} \tag{11.0}$$

a solution of which is available if s is a prime or a prime power.

If we take the s blocks formed by taking all possible combinations of $s - 1$ treatments from the ith group, we get an unreduced BIB design with parameters

$$v^* = b^* = s, \qquad r^* = k^* = s - 1, \qquad \lambda^* = s - 2. \tag{11.15}$$

Repeating this for each group and taking together all the BIB designs so formed we get the GD design

$$\begin{aligned} v &= s^2 - s, & b &= s^2 - s, & r &= s - 1, & k &= s - 1, \\ m &= s - 1, & n &= s, & \lambda_1 &= s - 2, & \lambda_2 &= 0. \end{aligned} \tag{11.2}$$

Taken by itself this is a disconnected design in the sense explained in [23] and [24], and any contrast between treatments of different groups is nones-

timable. But if we take together the blocks of (11.0) and (11.2) we get the GD design

$$
\begin{aligned}
&v = s^2 - s, \quad b = 2s^2 - s, \quad r = 2s - 1, \quad k = s - 1, \\
&m = s - 1, \quad n = s, \quad \lambda_1 = s - 2, \quad \lambda_2 = 1.
\end{aligned} \tag{11.35}
$$

As an illustration we give below the blocks for the case $s = 4$ (Design (3) of Table X).

$$
\begin{array}{cccccccccccccccccccccc}
1 & 4 & 3 & 2 & 2 & 4 & 1 & 3 & 4 & 2 & 1 & 3 & 1 & 2 & 3 & 4 & 2 & 1 & 1 & 1 & 6 & 5 & 5 & 5 & 10 & 9 & 9 & 9 \\
8 & 5 & 6 & 7 & 8 & 6 & 7 & 5 & 7 & 5 & 6 & 8 & 5 & 6 & 7 & 8 & 3 & 3 & 2 & 2 & 7 & 7 & 6 & 6 & 11 & 11 & 10 & 10 \\
11 & 9 & 12 & 10 & 9 & 10 & 12 & 11 & 11 & 12 & 9 & 10 & 10 & 11 & 9 & 12 & 4 & 4 & 4 & 3 & 8 & 8 & 8 & 7 & 12 & 12 & 12 & 11.
\end{array} \tag{11.4}
$$

The corresponding groups are given by the columns of the scheme

$$
\begin{array}{ccc}
1 & 5 & 9 \\
2 & 6 & 10 \\
3 & 7 & 11 \\
4 & 8 & 12.
\end{array} \tag{11.45}
$$

TABLE X

Parameters of GD designs obtainable by extension and addition

Serial no.	Parameters							
	v	b	r	k	m	n	λ_1	λ_2
(1)	12	24	6	3	6	2	2	1
(2)	24	54	9	4	8	3	3	1
(3)	12	28	7	3	3	4	2	1
(4)	20	45	9	4	4	5	3	1
(5)	12	32	8	3	2	6	2	1

The first 16 blocks of (11.4) are obtained by taking the first three rows of (7.5), whereas the remaining 12 blocks are obtained by taking all combinations of three treatments from each group.

By taking $s = 5$ in (11.35) we get design (4) of Table X.

(b) Suppose we have solutions available for GD designs with parameters

$$v = mn, \quad b, \quad r, \quad k, \quad m, \quad n, \quad \lambda_1, \quad \lambda_2 \tag{11.5}$$

$$v' = mn/\alpha, \quad b', \quad r', \quad k' = k, \quad m' = m/\alpha, \quad n' = n, \quad \lambda_1', \quad \lambda_2' \tag{11.6}$$

where m' and α are integers, and

$$\lambda_1 + \lambda_1' = \lambda_2 + \lambda_2' = \lambda_1'' \text{ (say)}. \tag{11.65}$$

The m groups of (11.5) can be divided into α sets each of m' groups. With the v' treatments occurring in any such set we can write down a solution for (11.6).

If we do this for each set and add the $\alpha b'$ blocks so obtained to the blocks of (11.5) we get the solution of a GD design with parameters

$$\begin{aligned} &v'' = mn, \quad b'' = b + b'\alpha, \quad r'' = r + r', \quad k'' = k, \\ &m'' = \alpha, \quad n'' = m'n, \quad \lambda_1'', \quad \lambda_2'' = \lambda_2 \end{aligned} \tag{11.7}$$

where the treatments occurring in a set now belong to the same group. Obviously every treatment occurs $r + r'$ times in the final design, but we have to show that any two treatments belonging to the same set occur together λ_1'' times, and any two treatments belonging to different sets occur together λ_2 times.

If two treatments belong to the same set, they either occur together in the same group or in different groups. In the first case they occur together in λ_1 blocks obtained from (11.5) and in λ_1' blocks obtained from (11.6). In the second case they occur together in λ_2 blocks obtained from (11.5) and λ_2' blocks obtained from (11.6). It follows from (11.65) that in either case they occur together λ_1'' times.

Again if two treatments belong to different sets they will occur together in λ_2 blocks obtained from (11.5) and in no blocks obtained from (11.6). This completes the proof.

As an illustration let us start with the GD design with parameters

$$\begin{aligned} &v = 12, \quad b = 20, \quad r = 5, \quad k = 3, \\ &m = 6, \quad n = 2, \quad \lambda_1 = 0, \quad \lambda_2 = 1, \end{aligned} \tag{11.75}$$

the blocks of which are given by the columns of (5.7) and the groups by (5.8).

Let us take $\alpha = 2$, and let the first three groups belong to the first set and the last three groups to the second set. Also as noted in Section 10 a solution of the GD design with parameters

$$\begin{aligned} &v' = b' = 6, \quad r' = k' = 3, \quad m' = 3, \\ &n' = 2, \quad \lambda_1' = 2, \quad \lambda_2' = 1 \end{aligned} \tag{11.8}$$

is given by the last six columns of (10.6), and the groups by the last three columns of (10.65). We note that $\lambda_1 + \lambda_1' = \lambda_2 + \lambda_2' = 2$. Hence we can build up a solution of

$$\begin{aligned} &v = 12, \quad b = 32, \quad r = 8, \quad k = 3, \\ &m = 2, \quad n = 6, \quad \lambda_1 = 2, \quad \lambda_2 = 1 \end{aligned} \tag{11.85}$$

by adding to the solution of (11.75) a solution of (11.8) twice over identifying $a_0, a_1, a_2, a_3, a_4, a_5$, once with 5, 11, 2, 7, 6, 8 and next with 9, 10, 4, 12, 1, 3 respectively. Thus the 32 blocks of (11.85) are given by the 20 columns of the scheme (5.7) together with the twelve columns of the following scheme

$$\begin{array}{cccccccccccc} 5 & 11 & 2 & 7 & 6 & 8 & 9 & 10 & 4 & 12 & 1 & 3 \\ 11 & 2 & 7 & 6 & 8 & 5 & 10 & 4 & 12 & 1 & 3 & 9 \\ 7 & 6 & 8 & 5 & 11 & 2 & 12 & 1 & 3 & 9 & 10 & 4. \end{array} \tag{11.9}$$

The first group consists of the treatments 5, 11, 2, 7, 6, 8 and the second group consists of 9, 10, 4, 12, 1, 3.

The parameters of GD designs obtainable by extension and addition are shown in Table X.

REFERENCES

[1] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced incomplete block designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[2] R. C. Bose and W. S. Connor, "Combinatorial properties of group divisible designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[3] R. C. Bose and K. R. Nair, "Partially balanced incomplete block designs," *Sankhya*, Vol. 4 (1939), pp. 337–372.

[4] K. R. Nair and C. R. Rao, "A note on partially balanced incomplete block designs," *Science and Culture*, Vol. 7 (1942), pp. 568–569.

[5] C. R. Rao, "General methods of analysis for incomplete block designs," *J. Amer. Stat. Assn.*, Vol. 42 (1947), pp. 541–561.

[6] F. Yates, "Incomplete randomized blocks," *Ann. Eugenics*, Vol. 7, (1936), pp. 121–140.

[7] R. A. Fisher and F. Yates, *Statistical Tables*, 3rd ed., Hafner Publishing Co., New York, 1948.

[8] R. C. Bose, "On the construction of balanced incomplete block designs," *Ann. Eugenics*, Vol. 9 (1939), pp. 353–399.

[9] R. C. Bose, "Some new series of balanced incomplete block designs," *Bull. Calcutta Math. Soc.*, Vol. 34 (1942), pp. 17–31.

[10] R. C. Bose, "A note on the resolvability of balanced incomplete block designs," *Sankhya*, Vol. 6, (1942), pp. 105–110.

[11] W. G. Cochran and G. M. Cox, *Experimental Designs*, John Wiley and Sons, 1950.

[12] R. C. Bose, "An affine analogue of Singer's theorem," *J. Indian Math. Soc.* (new series), Vol. 6 (1942), pp. 1–15.

[13] J. Singer, "A theorem in finite projective geometry and some applications to number theory," *Trans. Amer. Math. Soc.*, Vol. 43 (1938), pp. 377–385.

[14] C. R. Rao, "Difference sets and combinatorial arrangements derivable from finite geometries," *Proc. Nat. Inst. Sci. India*, Vol. 12 (1946), pp. 123–135.

[15] R. C. Bose, "On a resolvable series of balanced incomplete block designs," *Sankhya*, Vol. 8 (1947), pp. 249–256.

[16] C. R. Rao, "Hypercubes of strength d leading to confounded designs in factorial experiments," *Bull. Calcutta Math. Soc.*, Vol. 38 (1946), pp. 67–78.

[17] C. R. Rao, "Factorial experiments derivable from combinatorial arrangements of arrays," *J. Roy. Stat. Soc.*, Suppl. Vol. 9 (1947), pp. 128–139.

[18] C. R. Rao, "On a class of arrangements," *Proc. Edinburgh Math. Soc.*, Vol. 8 (1949), pp. 119–125.

[19] R. L. Plackett and J. P. Burman, "The design of optimum multifactorial experiments," *Biometrika*, Vol. 33 (1943–46), pp. 305–325.

[20] K. A. Bush, "A generalization of a theorem due to MacNeish," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 293–295.

[21] K. A. Bush, "Orthogonal arrays of index unity," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 426–434.

[22] R. C. Bose and K. A. Bush, "Orthogonal arrays of strength two and three," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 508–524.

[23] R. C. Bose, "The design of experiments: Presidential address to the section of Statistics, 34th Indian Science Congress." (1947).

[24] R. C. Bose, "Least square aspects of the analysis of variance," Institute of Statistics, University of North Carolina, Mimeo. Series 9.

THE BASIC THEOREMS OF INFORMATION THEORY

By Brockway McMillan

Bell Telephone Laboratories

Summary. This paper describes briefly the current mathematical models upon which communication theory is based, and presents in some detail an exposition and partial critique of C. E. Shannon's treatment of one such model. It then presents a general limit theorem in the theory of discrete stochastic processes, suggested by a result of Shannon's.

1. General models of the communication problem.

1.0. *Introduction.* For the purposes of this exposition, information theory is the body of statistical mathematics which has developed, largely over the last decade, out of efforts to understand and improve the communications art. We shall not attempt a history of this development, nor any detailed justification for its existence, since either of these efforts would take us further into the technics of communication than is desirable in a short essay.

It suffices to say here that this discipline has come specifically to the attention of mathematicians and mathematical statisticians almost exclusively through the book [1] of N. Wiener and the paper [2] of C. E. Shannon.

In the remainder of this section we shall describe very broadly the kind of problem to which these two works are addressed.

1.1. *A simple model.* The simplest mathematical model of the communication problem is like the problem of parameter estimation. A parameter θ, usually ranging over a fairly abstract or at least multi-dimensional domain, represents the transmitted message. A variable, y, also fairly abstract in general, represents the received message. In realistic situations the received message is seldom a mathematically exact copy, or even an exactly predictable mutilation, of the original transmitted message. Hence, y is represented as a random variable whose distribution depends upon the parameter θ. The communication problem then is: given a sample of one value of y, to estimate the unknown θ.

There are two reasons why this model may not seem at first look to be a good one for the communication problem. One is merely that our most usual media of communication, direct acoustic transmission of voice and the written or printed word, are ones in which essentially exact transmission is possible and we are not aware of the underlying statistical nature of the problem. This is clearly a matter of degree, however, and almost anyone can find in his own experience instances in which the statistical aspect of the problem was evident.

Another apparent failing of this model is in fact real, and has led to refinement of the model. There are communication problems, mostly in technical fields, where it is realistic to assume that the recipient of y has no a priori knowledge

Received 8/9/51.

about the parameter θ. The usual situation in human experience, however, is one in which there is a great deal of a priori knowledge about the possible values of θ. There are simple experiments with mutilated text, spoken or written, which will convince one that he can, and often does, exploit his own a priori knowledge of language, speaker, and subject matter to assist in deciphering what he reads and hears. A realistic model must include this possibility.

1.2. *Stochastic transmitted message.* It was Wiener who first clearly pointed out that we may, and indeed often must, regard the transmitted message itself as a random variable drawn from a universe whose distribution function reflects our a priori knowledge of the situation. Cogent statements of this philosophy may be found both in [1] and [2]. This leads us to a model in which we have two abstract random variables, say x representing the transmitted message (replacing the parameter θ), and y the received message. There is then a joint distribution function for x and y which contains in it the complete mathematical description of the situation. One ordinarily thinks of this distribution function as being "factored" into an a priori distribution for x, representing the universe of possible messages, and a conditional distribution for y knowing x, representing for each x the universe of possible mutilations thereof.

In this second and more important model, one can still regard the communication problem as one of estimation: given the y value of a joint sample (x, y), to estimate the x value. This view is particularly appropriate in discussing the work of [1]. Here, the x and y are numerically valued time series and there is a natural numerical way to measure the deviation between the estimated and true values of x, namely, by the variance of estimate.

The statistician may alternatively wish to regard the communication problem (in either model) as one of testing hypotheses. The observed y has a distribution depending on the hypothesis "x;" the problem is to decide which x is obtained at the time of observation. This view is more appropriate to the work of [2], wherein the time series are abstract valued, and no natural measure of the "wrongness" of an incorrectly adopted hypothesis is available. In the second model, the a priori distribution for hypotheses x eliminates one kind of testing error, so that in this model there is a simple criterion of performance, namely, the total probability in the (x, y) universe of all events (x, y) in which the hypothesis adopted is correct. The reader will observe this particular criterion in sections 6 and 8.

The distinction between estimation, on the one hand, and testing among many hypotheses, on the other, is not sharp. We shall use "estimation" as a loose word to refer to the kind of model here set up for communication.

1.3. *Peculiarities of engineering applications.* Information theory is distinguished from a general study of models like these in two important respects. In the first place, as noted, the random quantities x and y of interest are, naturally, time series. Furthermore, the passage of time is explicitly recognized and the distinction between past events, which can be known, and future events which cannot be known, is carefully observed.

In the second place, the kind of question considered in information theory, particularly by Shannon, reflects the peculiar interests of communication engineers. To illustrate this, we might go back to the jointly distributed abstract variables x and y of 1.2 above and the estimation problem there stated: given a sample of one value of y, to estimate the corresponding x. Typically, a practicing statistician facing this kind of problem will find himself confronted with a given joint distribution function for the variables, or at least committed to choosing one which he thinks is representative, and his attention is directed toward such questions as the following.

a. By what criterion shall various estimates of x be compared?

b. Given the criterion, what is the best estimate of x which can be made, and how good is it?

c. How do competing methods of estimating x compare with the best?

These questions of course appear in a communication context, too. The entire effort of [1] is concentrated in this general area. It often happens, however, that the communication engineer has a freedom that the statistician seldom has, that of controlling, at least in part, the joint distribution with which he must deal. How this comes about will be discussed in a moment. We can see at once, however, that his interest in question (b) above will then extend to asking, in addition, how he can optimize his best estimate over the additional freedom he has.

1.4. *The additional freedom.* The additional freedom enjoyed by a communication engineer is like the freedom granted the designer of an experiment. Typically, technology provides the engineer with a communicating device or medium; a random variable y whose range Y represents, as above, the events which can take place at the receiving point, and a probability distribution for y which depends upon a parameter θ. As above, the range Θ of θ represents the possible events at the transmitting point. In addition, one is given a quite separate random variable x whose range X is the universe of possible messages with a probability measure appropriate thereto.

No relation is yet specified between the message x and the "stimulus" θ which is applied to the communication medium, and it is here that the extra freedom lies. Subject to limitations set by the necessary distinction between past and future, one is free to choose a mapping function $f(x)$ from X into Θ, $\theta = f(x)$. This corresponds to choosing some kind of encoding or modulation scheme transforming the original message into a form suitable for transmission.

To illustrate the effect of this, suppose that the distribution function of y has a density $\rho(\theta; y)$ with respect to some fixed underlying measure ν in the y universe, and that the distribution of x has a density $\sigma(x)$ with respect to some underlying measure μ in the x universe. Then if one fixes the relation above between x and θ, the function $\sigma(x)\rho(f(x); y)$ in $X \otimes Y$ represents the density of the resulting joint distribution of x and y relative to the product measure $\mu \otimes \nu$. It is this joint distribution with which the communication engineer works.

To the practising engineer, the most interesting theorems of Shannon's paper relate to what can be achieved by varying the encoding process repre-

sented by the function $f(x)$. The strong theorems now known are all of an asymptotic kind.

1.5. *Role of Fourier analysis.* Even the casual reader will observe in [1], and in the latter part of [2], a preoccupation with Fourier analysis. It may be well to point out that this is a kind of accident; it happens that most practical communication media are governed by linear time-invariant differential equations. Hence, the first applications of information theory have been to systems which are naturally best handled by the tools of Fourier or Laplace analysis.

2. Terminology and concepts.

2.0. *Limitation to discrete model.* We shall confine our attention to the first part of Shannon's paper [2]. This whole paper relates to the second model of the communication problem described above, with an emphasis on the kind of question discussed in 1.3. The first part of that paper is based on a fairly specific kind of model. The stochastic processes which it admits are all derived from Markov processes having finitely many states. The auxiliary devices, encoders, etc., which are admitted are defined by similar constructions. We adopt the term "finitary" to denote a restriction to these classes of objects without at this point repeating Shannon's definitions in detail. (There is a restriction, tacit in [2] but nowhere made explicitly, to devices whose graphs have the property that the terminal state of any transition is uniquely fixed when the initial state and the letter emitted are given. For the present, we take "finitary" to include this limitation.)

The central concepts of [2] may be introduced well enough here by a glossary of terms. At this purely descriptive level, we may be quite general and admit things which are not finitary.

2.1. *Sample space and measurable sets.* Let A be a finite set. We call such a set an alphabet and will have occasion to introduce further alphabets A_1, B, etc. These are all abstract finite sets. An element of A will be called a letter of A, or simply a letter when no ambiguity results.

Let I denote the set of integers: $I = (\cdots, -1, 0, 1, 2, \cdots)$.

Given an alphabet A, denote by A^I the class of infinite sequences

$$x = (\cdots, x_{-1}, x_0, x_1, x_2, \cdots)$$

where each $x_t \varepsilon A$, $t \varepsilon I$. Here x is an element of A^I, and we call x_t the letter of x at time t.

A basic set (in A^I) is a subset of A^I obtained by specifying

(i) an integer $n \geqq 1$,

(ii) a finite sequence $\alpha_0, \alpha_1, \cdots, \alpha_{n-1}$ of letters $\alpha_k \varepsilon A$.

(iii) an integer t, $-\infty < t < \infty$.

The basic set resulting from this specification consists of all sequences $x \varepsilon A^I$ such that

$$x_{t+k} = \alpha_k, \qquad 0 \leqq k \leqq n - 1.$$

Let F_A be the Borel field of subsets of A^I determined by the basic sets.

2.2. *Glossary.* Our glossary now reads:

2.21. *Information source.* If μ is a probability measure defined over the Borel field F_A, the ensemble or stochastic process $[A^I, F_A, \mu]$ is an information source. Since the space A^I is fixed by the alphabet, and the Borel field F_A is always that determined by the basic sets, we can specify a source by the pair of symbols $[A, \mu]$.

2.22. *Stationary and ergodic sources.* Consider a source $[A, \mu]$. Let T be the coordinate-shift transformation defined as follows. If $x = (\cdots, x_{-1}, x_0, x_1, \cdots)$ then $Tx = (\cdots, x'_{-1}, x'_0, x'_1, \cdots)$, where $x'_t = x_{t+1}$, $t \varepsilon I$. Then T preserves membership in F_A (measurability). The source will be called stationary if (i) below holds, and ergodic if (i) and (ii) both hold.

(i) If $S \varepsilon F_A$, then $\mu(S) = \mu(TS)$.

(ii) If $S = TS$, then either $\mu(S) = 0$ or $\mu(S) = 1$.

2.23. *Transducer.* A transducer is characterized by two alphabets, A and B, and a function τ from A^I to B^I: given $x \varepsilon A^I$, $\tau(x) \varepsilon B^I$. A transducer differs from a general functional relationship in that it cannot anticipate.

If $x^{(1)} \varepsilon A^I$ and $x^{(2)} \varepsilon A^I$ and t_0 is an integer such that

$$x_t^{(1)} = x_t^{(2)} \qquad \text{for } t \leqq t_0,$$

then

$$y_t^{(1)} = y_t^{(2)} \qquad \text{for } t \leqq t_0,$$

where

$$y^{(i)} = \tau(x^{(i)}), \qquad i = 1, 2.$$

We can specify a transducer by the symbol $[A, \tau, B]$.

2.24. *Channel or communication channel.* A channel is characterized by two alphabets A and B, and a list of probability measures ν_θ defined over F_B, one for each $\theta \varepsilon A^I$. Here we have used θ to denote the "parameter" in conformance with an earlier notation.

Like a transducer, a channel cannot anticipate. That is, informally, if

$$\theta_t^{(1)} = \theta_t^{(2)} \qquad \text{for } t \leqq t_0, \tag{1}$$

we must have

$$\nu_1(S) = \nu_2(S), \tag{2}$$

where $\nu_i(S)$ denotes the value of $\nu_\theta(S)$ when $\theta = \theta^{(i)}$, for any set $S \varepsilon F_B$ which depends only on letters occuring before $t_0 + 1$. More precisely stated, (1) must imply (2) for any set $S \varepsilon F_B$ such that "$y_t^{(1)} = y_t^{(2)}$ for $t \leqq t_0$, and $y^{(1)} \varepsilon S$" implies "$y^{(2)} \varepsilon S$."

A transducer is a special case of a channel; it is a channel in which the received signal y is determined exactly by the transmitted signal θ.

We can specify a channel by the symbol $[A, \nu_\theta, B]$.

2.25. *Stationarity.* The concept of stationarity extends to channels and transducers. It suffices to define a stationary channel, a stationary transducer is a special case. Referring to the definition of a channel, this channel will be called stationary if, for any $S \varepsilon F_B$, $\nu_\theta(S) = \nu_{T\theta}(TS)$, where T is the coordinate-shift transformation.

2.26. We have so worded the definitions above that all sources are "letter generators" producing one new letter for each unit of time, and channels and transducers accept and produce one letter for each unit of time. In a careful setting of the theory, one must account for the phenomena of compression and expansion which appear when languages are translated. For example, a long business message of fairly stereotyped form, when encoded for transmission by cable, may appear in a form having many fewer letters or words than the original. There are several ways of accommodating the mathematics to this situation, but these details are unimportant in a first look at the subject and will be ignored from here on. The fact of so ignoring them does not invalidate any theorem that will be stated. It merely leaves a gap between these theorems and certain useful interpretations of them.

3. Entropy.

3.0. *Entropy.* The terms defined in Section 2, suitably hedged, are the concepts with which [2] deals. (For purposes of exposition, we have defined channels and transducers quite differently from [2]. The disparity is largely but not entirely verbal. (Cf. 10.2.)) The principal tool for their quantitative study is the concept of entropy.

Let $p_1, p_2, \cdots, p_n$ be a finite and exhaustive list of probabilities: $p_i \geqq 0$, $1 \leqq i \leqq n$, $p_1 + p_2 + \cdots + p_n = 1$. The entropy of this list is defined to be

$$H(p_1, p_2, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log p_i = \text{Expectation } (-\log p).$$

It is by now traditional to use logs to the base 2 in this definition, but the choice of base affects the value of H only by a constant factor. We shall use the base 2.

3.1. *Marginal entropies.* To change the notation slightly, suppose that α and β run over finite index sets (alphabets) A and B, and that $p(\alpha, \beta)$ is the probability of the joint event (α, β). That is $p(\alpha, \beta) \geqq 0$, $\sum_{\alpha \varepsilon A} \sum_{\beta \varepsilon B} p(\alpha, \beta) = 1$. The entropy of this list of probabilities is denoted by $H(\alpha, \beta)$:

$$H(\alpha, \beta) = -\sum_{\alpha} \sum_{\beta} p(\alpha, \beta) \log p(\alpha, \beta).$$

We can define also two marginal entropies

$$H(\alpha) = -\sum_{\alpha} \sum_{\beta_1} p(\alpha, \beta_1) \log \left(\sum_{\beta} p(\alpha, \beta)\right),$$

$$H(\beta) = -\sum_{\beta} \sum_{\alpha_1} p(\alpha_1, \beta) \log \left(\sum_{\alpha} p(\alpha, \beta)\right),$$

and two average conditional entropies

$$H_\beta(\alpha) = H(\alpha, \beta) - H(\beta),$$

$$H_\alpha(\beta) = H(\alpha, \beta) - H(\alpha).$$

3.2. *Average conditional entropy.* These latter are called average conditional entropies because of the following formula: fix β and consider the conditional probabilities for the various $\alpha \varepsilon A$. These are $q_\beta(\alpha) = p(\alpha, \beta)/\sum_{\alpha_1} p(\alpha_1, \beta)$. The entropy of this list is

$$-\sum_\alpha q_\beta(\alpha) \log q_\beta(\alpha) = \frac{1}{r(\beta)} \sum_\alpha p(\alpha, \beta) \log p(\alpha, \beta) + \frac{1}{r(\beta)} \sum_\alpha p(\alpha, \beta) \log \sum_{\alpha_2} p(\alpha_2, \beta) \tag{1}$$

where $r(\beta)$ is defined by (2) below.

This expression is the entropy of the conditional distribution of α when it is known that a particular β has occurred. The a priori probability of this β is

$$\sum_{\alpha_1} p(\alpha_1, \beta) = r(\beta). \tag{2}$$

To average (1) over all β, we multiply it by (2) and sum over β. The result is seen to be $H_\beta(\alpha)$. This last entropy, then, is the average over all β of the entropies of the conditional distribution of α when β is known.

3.3. *Properties.* Shannon [2] gives a fairly complete heuristic justification for regarding the entropy of a list of probabilities as a measure of one's a priori uncertainty as to which of the possible events will actually occur in a given trial. In the course of this demonstration, he introduces the most important mathematical properties of the H function. These are (i) its positivity, (ii) a kind of convexity property implied by the convexity of the function $-x \log x$, (iii) that composition law which permitted the identification above of the average value of (1) over β, with the earlier defined $H_\beta(\alpha)$, and (iv) $H = 0$ if and only if there is exactly one event of nonzero probability.

The convexity property (ii) mentioned above leads to the general inequality $H_\beta(\alpha) \leqq H(\alpha)$; that is, verbally, a condition (i.e. an a priori restriction on the "freedom of choice") never increases an entropy. This statement must however be taken only in the average sense in which it is stated: for any particular β, the entropy of the conditional distribution of α bears no provable relation to the marginal entropy $H(\alpha)$. It is only in the average over-all β that an inequality obtains.

4. The entropy rate of a source.

4.0. *Definition.* So far we have considered the entropy of a list of probabilities. The entropy rate of a stationary source $[A, \mu]$ is most easily defined as follows. Given $x \varepsilon A^I$, we use either of the bracket notations

(1) $$[x_t, x_{t+1}, \cdots, x_{t+n-1}], [t, t+n-1; x]$$

to denote that basic set $S \subseteq A^I$ which consists of all x' such that

$$x'_{t+h} = x_{t+h}, \qquad 0 \leqq h \leqq n-1.$$

The second notation will be used when it is to be emphasized that the basic set depends upon a particular infinite sequence x.

The possible basic sets (1), as x ranges over A^I, or, alternatively, as the x_{t+h}, $0 \leqq h \leqq n-1$, range independently over A, partition A^I into a^n measurable subsets, where a is the number of letters in the alphabet A. These subsets represent all the possible sequences of n consecutive letters. They have the respective probabilities

(2) $$\mu([x_t, x_{t+1}, \cdots, x_{t+n-1}]).$$

Our stationarity assumption makes this list of probabilities independent of t. There is then a unique number F_n, independent of t, which is the entropy of this list (2) of probabilities. We shall show presently that the limit

(3) $$\lim_{n\to\infty} \frac{1}{n} F_n$$

always exists. The value of this limit is defined to be the entropy rate of the source $[A, \mu]$.

4.1. *Interpretation.* One cannot escape the heuristic meaning of this rate; one considers the possible long sequences of text as his universe of events, and evaluates the uncertainty F_n of the outcome of a trial. This uncertainty is then prorated among the n letters. These letters represent interdependent but possibly not determinately related elementary events whose concatenation generates the universe. The result, F_n/n, represents in the limit the average uncertainty per letter generated by the source.

4.2. *Defining F_n as an integral.* We shall now prove the existence of the limit (3). The proof follows Shannon's in a different notation.

Given any $x \in A^I$, the basic set (1) defined by that x contains x. The probability (2) then may be regarded as a step function of x, equal for each x to the probability of that basic set containing x which is specified by letter values at times $t, t+1, \cdots, t+n-1$. In the same way, the definition

(4) $$f_n(x) = -\frac{1}{n} \log \mu([0, n-1; x])$$

defines a nonnegative step function of x. One verifies at once from the definition of F_n that

(5) $$\frac{1}{n} F_n = \int_{A^I} f_n(x)\, d\mu(x).$$

Regarding (2) and (4) as functions of x in this way permits us to phrase certain key problems in the language of integration theory.

4.3. *Another definition of* H. Consider now the special conditional probabilities

$$p_n(x) = \frac{\mu([x_{-n}, x_{-n+1}, \cdots, x_{-1}, x_0])}{\mu([x_{-n}, \cdots, x_{-1}])}, \qquad n \geqq 1. \tag{6}$$

Again we use the device of representing these as step functions of x. In words, $p_n(x)$ is the conditional probability of observing at time zero the letter x_0 of x, when it is known that the letters occurring at times $t = -n, -n+1, \cdots, -1$ are exactly those of x.

Define

$$\begin{aligned} g_0(x) &= f_1(x) \\ g_n(x) &= -\log p_n(x), \qquad n \geqq 1. \end{aligned} \tag{7}$$

Then $g_n(x) \geqq 0$.

One verifies by direct calculation from (6) and (7) that

$$G_n = \int_{A^I} g_n(x)\, d\mu(x) \tag{8}$$

is the average conditional entropy of the next letter when n preceding letters are known. The inequality stated earlier, that adjoining a condition cannot increase an entropy, can be used to show that the G_n form a monotone sequence:

$$G_0 \geqq G_1 \geqq G_2 \geqq \cdots \geqq 0.$$

Therefore

$$\lim_{n \to \infty} G_n = H \tag{9}$$

certainly exists. The verbal interpretation of G_n, the average conditional entropy of the next letter after a long segment of text is already known, suggests that the limit H in (9) is again the average uncertainty per letter generated by the source, that is, H is the entropy rate defined in (3). The proof in 4.4 below that this is indeed so, proves the existence of the limit (3).

4.4. *Identification of two definitions.* By a direct calculation from the definitions it is found that

$$f_N(x) = \frac{1}{N} \sum_{k=0}^{N-1} g_k(T^k x). \tag{10}$$

If one integrates this and uses the assumed stationarity of μ, he obtains

$$\frac{1}{N} F_N = \frac{1}{N} (G_0 + G_1 + \cdots + G_N). \tag{11}$$

Therefore F_N/N represents the first Cesaro mean of a monotonely convergent sequence. It follows that the limit (3) exists and indeed is approached monotonely. A further consequence of (11) is that $F_N/N \geqq G_N \geqq H$.

5. The capacity of a channel.

5.0. *Channel and source.* We wish now to examine a stationary channel "driven" by a stationary source. Consider a source $[A, \mu]$, and a channel $[A, \nu_\theta, B]$. Denote by $C = A \otimes B$ the alphabet of pairs (α, β), $\alpha \varepsilon A$, $\beta \varepsilon B$. Then C^I is the class of all infinite sequences

$$(\cdots, (x_{-1}, y_{-1}), (x_0, y_0), (x_1, y_1), \cdots)$$

where $x_t \varepsilon A$, $y_t \varepsilon B$, $t \varepsilon I$. In an obvious way we can also regard $C^I = A^I \otimes B^I$, that is, as the class of paired sequences (x, y), $x \varepsilon A^I$, $y \varepsilon B^I$. It is known that the Borel field F_C is determined by the sets $X \otimes Y$ where $X \varepsilon F_A$, $Y \varepsilon F_B$. We define a measure ω for sets in F_C by the formula

$$\int_{C^I} h(x, y)\, d\omega(x, y) = \int_{A^I} d\mu(x) \int_{B^I} h(x, y)\, d\nu_x(y)$$

valid for all positive measurable $h(x, y)$. Here ν_x is the measure over F_B which is induced by the channel when the input sequence is $x \varepsilon A^I$.

The stochastic process $[C^I, F_C, \omega]$ is now a source, which we denote by $[C, \omega]$.

It is easily shown that if the original source $[A, \mu]$ and the channel are stationary, then the source $[C, \omega]$ is stationary.

5.1. *Marginal distributions.* The source $[C, \omega]$ represents the joint distribution of x and y, of input to and output from the channel. The source $[A, \mu]$ represents the marginal distribution of the input. The marginal distribution of the output is represented by the source $[B, \eta]$, where the measure η over F_B is defined by

$$\int_{B^I} k(y)\, d\eta(y) = \int_{A^I} d\mu(x) \int_{B^I} k(y)\, d\nu_x(y).$$

This marginal source is staionary if $[A, \mu]$ and $[A, \nu_\theta, B]$ are.

5.2. *Causation.* It is worth noting that the implication of causation in our language here, as we speak of a channel driven by a source, results from the fact that we consider the channel $[A, \nu_\theta, B]$ as a pregiven thing, existing independently of any particular source $[A, \mu]$; this is the typical situation in the communications art. Actually, the joint process $[C, \omega]$ is a completely symmetrical concept, as to the roles of x and y, and one may consider, at will, the conditional probabilities $\nu_x(S)$, $x \varepsilon A^I$, $S \varepsilon F_B$, the conditional probabilities of y-events, knowing x, or the conditional probabilities, say, $\bar{\mu}_y(U)$, $y \varepsilon B^I$, $U \varepsilon F_A$, of x-events, knowing y. (Indeed, given the joint process, one will find that each of these conditional probabilities ν_x, respectively $\bar{\mu}_y$, are measures for, respectively, almost all $x(\mu)$, almost all $y(\eta)$.)

It happens that in most applications the ν_x are pregiven, and the $\bar{\mu}_y$ derivative.

5.3. *Channel capacity.* To use Shannon's notation, let $H(x, y)$ denote the entropy rate of the source $[C, \omega]$, $H(x)$ the entropy rate of the marginal source $[A, \mu]$, and $H(y)$ that of the marginal source $[B, \eta]$. The quantity $R = H(x) + H(y) - H(x, y)$ is defined to be the transmission rate achieved by the source

$[A, \mu]$ over the channel $[A, \nu_\theta, B]$. The supremum or least upper bound of these rates, as μ is allowed to vary, is defined to be the capacity of that channel.

5.4. *Interpretation.* An intuitive interpretation of the rate $H(x) + H(y) - H(x, y)$ can be obtained if we assume that the quantity $H_x(y) = H(x, y) - H(x)$ can be given the same verbal interpretation when the x and y are stochastic processes that it was given earlier when the random quantities involved were drawn from finite populations. That it can, in the same limiting sense that the entropy concept has been carried over to stochastic processes, is easy to show. Foregoing this demonstration, we observe that $R = H(y) - H_x(y)$; that is, the rate of transmission R is the marginal rate of the output, $H(y)$, diminished by that amount of uncertainty at the output which arises from the average uncertainty of y even when x is known, that is, by $H_x(y)$, the average conditional entropy of y when x is known. In this verbal way, at least, R represents that portion of the "randomness" or average uncertainty of each output letter which is not assignable to the randomness created by the channel itself.

Another observation here is also pertinent. Because of the symmetry of R in x and y (which is more than a mere consequence of the notation!) we also have $R = H(x) - H_y(x)$. This shows R as the rate of the original source diminished by the average uncertainty as to the input x when the output y is known.

6. The fundamental theorem.

6.0. *As justifying the theory.* So far, we have introduced a list of what is hoped are natural-seeming concepts, and have stated a few mathematical results to help justify the rather picturesque language used in introducing them. The concepts themselves can only be justified as objects worthy of mathematical attention by the existence of theorems relating them. There is one such theorem, the so-called fundamental theorem for a noisy channel ([2], Theorem 11), which in itself performs this task completely. We shall quote this theorem and sketch its proof. This will complete our general exposition and lead us to our general limit theorem.

6.1. The Theorem. The fundamental theorem relates to this question. Suppose we are given a stationary channel with input alphabet A, and a stationary ergodic source with alphabet A_1. We are permitted to insert a stationary transducer $[A_1, \tau, A]$ between the source and channel, to create in effect, a new stationary channel with input alphabet A_1. With this freedom, what is the optimum transmission rate which can be achieved between source and output?

For the class of finitary sources, channels, and transducers, admitted in the model used in [2], this question is answered by Shannon's theorem: Let the given channel have capacity C and the given source have rate H. Then if $H < C$, for any $\epsilon > 0$ there exists a transducer such that a rate $R > H - \epsilon$ can be achieved. If $H \geqq C$, there exists similarly a transducer such that $C \geqq R > C - \epsilon$. No rate greater than C can be achieved.

Actually, Shannon's proof of this theorem proves the following more complete result.

THEOREM. *Let the given channel have capacity C and the given source have rate H. If $H < C$, then, given any $\epsilon > 0$, there exists an integer $n(\epsilon)$ and a transducer (depending on ϵ) such that when $n(\epsilon)$ consecutive received letters are known, the corresponding n transmitted letters can be identified correctly with probability at least $1 - \epsilon$. If $H > C$ no such transducer exists.*

This statement is perhaps more satisfying to a statistician, in that the logarithmic quantities H and C appear only in the hypotheses. The conclusion is then given in terms of the criterion of performance suggested in 1.2.

6.2. *Interpretation.* In the vernacular, this theorem asserts that if a channel has adequate capacity C, an infinitesimal margin being mathematically adequate, then virtually perfect transmission of the material from the source can be achieved, but not otherwise. Here, of course, we have used "virtually perfect" to describe transmission at a rate

$$R = H - \epsilon_1 \geqq H - \epsilon. \tag{1}$$

The sense in which this is to be interpreted as virtually perfect transmission is, of course, an asymptotic one and refers to the rate at which certain probabilities decay as the amount of available received text increases.

Engineering experience has been that the presence in the channel of perturbations, noise, in the engineer's language, always degrades the exactitude of transmission. Our verbal interpretation above leads us to expect that this need not always be the case; that perfect transmission can sometimes be achieved in spite of noise. This practical conclusion runs so counter to naive experience that it has been publicly challenged on occasion. What is overlooked by the challengers is, of course, that "perfect transmission" is here defined quantitatively in terms of the capabilities of the channel or medium, perfection can be possible only when transmission proceeds at a slow enough rate. When it is pointed out that merely by repeating each message sufficiently often one can achieve virtually perfect transmission at a very slow rate, the challenger usually withdraws. In doing so, however, he is again misled, for in most cases the device of repeating messages for accuracy does not by any means exploit the actual capacity of the channel.

Historically, engineers have always faced the problem of *bulk* in their messages, that is, the problem of transmitting rapidly or efficiently in order to make a given facility as useful as possible. The problem of noise has also plagued them, and in many contexts it was realized that some kind of exchange was possible, for example, noise could be eliminated by slower or less "efficient" transmission. Shannon's theorem has given a general and precise statement of the asymptotic manner in which this exchange takes place.

The statistician will recognize the exchange between bulk and noise as akin to the more or less general exchange between sample size and validity or significance.

7. The asymptotic equipartition property.

7.0. *A Basic Lemma.* The theorem quoted in 6.1 is termed fundamental in

[2] because it answers a question which is clearly fundamental in the communications art, and because it defines the applicability of the central concept of channel capacity. Many of the later results in [2] then concern the calculation of capacities for practical or interesting channels.

The proof of this fundamental theorem rests directly on a lemma (Theorem 3 of [2]) which itself is a basic limit theorem in the theory of stochastic processes. As a mathematical theorem, this lemma requires very little of the specialized imagery of communication theory for its understanding. A mathematician, therefore, is likely to regard it as the more fundamental element. A generalization of it is the one contribution of the present paper.

7.1. *Shannon's form.* The basic limit theorem, as given in Theorem 3 of [2], asserts that the text from an ergodic finitary source possesses what we shall call an asymptotic equipartition property. The basic sets

$$[x_0, x_1, \cdots, x_{n-1}], \tag{1}$$

as x ranges over A^I describe a partition of A^I, as we noted earlier: a partition into a^n events, each one of which is the occurrence of a particular string of n letters. Shannon's Theorem 3 asserts that, if H is the rate of a finitary ergodic source, then, given $\epsilon > 0$ and $\delta > 0$, there exists an $n_0(\epsilon, \delta)$ such that, given any $n \geqq n_0$, the basic sets (1) above can be divided into two classes:

(i) a class whose union has μ-measure less than ϵ,

(ii) a class each member E of which has a measure $\mu(E)$ such that $| H + 1/n \log \mu(E) | < \delta$.

That is, this theorem asserts the possibility of dividing the long segments of text from a finitary source into a class of roughly equally probable segments plus a residual class of small total probability.

7.2. *Stronger form.* Let us introduce here the step functions $f_n(x)$ defined in (4) of Section 4: $f_n(x) = -1/n \log \mu([0, n-1; x])$. In terms of these, the possibility of dividing the long segments of text into the categories (i) and (ii) above is easily seen to be equivalent to the assertion that the sequence $(f_n(x))$ converges in probability to the constant H.

We shall say that a source $[A, \mu]$ has the asymptotic equipartition property, AEP, if the sequence $(f_n(x))$ converges in probability to a constant.

Shannon's Theorem 3 then asserts that a finitary ergodic source has the AEP. We shall improve this in Section 9 to read as follows.

THEOREM. *For any source $[A, \mu]$, the sequence $(f_n(x))$ converges in L^1 mean (μ). If $[A, \mu]$ is ergodic, and has rate H, this sequence converges in L^1 mean to the constant H.*

Since L^1 convergence here implies convergence in probability, (a fact easily proved,) we have the

COROLLARY. *Every ergodic source has the AEP.*

These are the limit theorems mentioned in the Summary. As we shall see in Section 8, they permit extending Shannon's fundamental theorem, 6.1, to other than finitary sources.

7.3. *Interpretation.* Returning to 7.1 and the description there of the AEP, we see that most of the probability must be accounted for by the aggregate (ii) of "likely" long sequences. That is, if the source has the AEP, there are, for large enough n, 2^{nH} likely basic sets $[x_0, \cdots, x_{n-1}]$, roughly equally probable, accounting in the aggregate for all but a small fraction of the total probability.

7.4. *Another corollary.* The proof in [2] of the fundamental theorem uses also a consequence of the AE property. We examine this consequence briefly.

Consider a stationary source $[A, \mu]$ and a stationary channel $[A, \nu_\theta, B]$. Suppose that both $[A, \mu]$, and the joint process $[C, \omega]$ which results when this source drives the channel, have the AEP. We can write

$$
\begin{aligned}
&-\frac{1}{n}\log \omega([0, n-1; x] \otimes [0, n-1; y]) \\
&\quad = -\frac{1}{n}\log \frac{\omega([0, n-1; x) \otimes [0, n-1; y)]}{\mu([0, n-1; x])} - \frac{1}{n}\log \mu([0, n-1; x]).
\end{aligned} \tag{2}
$$

Our hypothesis that the joint process has the AEP now implies that the left member of this equation converges in measure to a constant, namely the entropy rate of the joint process, $H(x, y)$. (Here the notation is misleading. In $H(x, y)$, the x and y are labels merely. Equation (2) involves x and y as specific variables.) Also by hypothesis the second term on the right converges in probability to $H(x)$, the entropy rate of $[A, \mu]$. It follows then that the first term on the right converges in probability also to a constant, which constant must then be $H_x(y)$, by 5.4.

Now the first term on the right of (2) is

$$
-\frac{1}{n}\log \bar{\mu}_{n,x}([0, n-1; y]), \tag{3}
$$

where the argument of the logarithm is the conditional probability of $[0, n-1; y]$ knowing that $[0, n-1; x]$ has occurred. We have therefore proved the following.

COROLLARY. *If* $[A, \mu]$ *and* $[C, \omega]$ *have the AEP, then the functions* (3) *converge in probability to a constant.*

8. Proof of the fundamental theorem.

8.0. *Introduction.* For simplicity, we do not examine the question of ergodicity and consider only the most interesting of the cases cited in the statement of the theorem (6.1), that in which we are given a finitary source $[A_1, \mu]$ of entropy rate H and a finitary channel $[A, \nu_\theta, B]$ of capacity $C > H$, both stationary. Our problem is then, given $\epsilon > 0$, to exhibit an $n(\epsilon)$ and a finitary transducer $[A_1, \tau, A]$ such that, when the given source drives the channel through this transducer it is possible at the receiver, given $n(\epsilon)$ consecutive received letters, to identify the corresponding $n(\epsilon)$ transmitted letters correctly with a probability exceeding $1 - \epsilon$. Here the probability is not conditional (i.e., not given the received letters) but in the universe of joint events at transmitter and receiver.

We shall review Shannon's argument. He does not supply detailed epsilontics here, and we shall not either. Generally, the manner in which they could be supplied is evident enough, though at one point we must consider a detail. (My efforts to make them simple have so far failed, however.)

8.1. *The "likely" events.* The channel $[A, \nu_\theta, B]$ has capacity $C = H + 2\gamma$, say, where $\gamma > 0$. There, therefore, exists a source, say $[A, \mu^*]$, which achieves over this channel a rate

$$R^* \geqq C - \gamma = H + \gamma. \tag{1}$$

We will use asterisks to denote quantities referring to this source. Let H^* be the entropy rate of $[A, \mu^*]$, and let $[C, \omega^*]$ denote the joint process of input (x) and output (y) when $[A, \mu^*]$ drives the channel. For a simpler notation let $K^* = H_y^*(x)$, the average conditional entropy of input to $[C, \omega^*]$ when output is known. Then by definition (5.3)

$$R^* = H^* - K^*. \tag{2}$$

We now invoke the AEP for the processes $[A_1, \mu]$, $[A, \mu^*]$, and $[C, \omega^*]$. For large n there are roughly 2^{nH} equally likely basic sets $[0, n-1; w]$ from $[A_1, \mu]$, call these the likely outputs of $[A_1, \mu]$. Similarly there are roughly 2^{nH^*} equally likely basic sets $[0, n-1; x]$ from $[A, \mu^*]$, the likely outputs of $[A, \mu^*]$. Furthermore, consider the possible basic sets $[0, n-1; y]$ at the output of the channel. With the exception of an aggregate of these of small total probability in $[C, \omega^*]$, the conditional probabilities in $[C, \omega^*]$ of the $[0, n-1; x]$, knowing $[0, n-1; y]$, are such that roughly, there are 2^{nK^*} equally likely $[0, n-1; x]$ for each $[0, n-1; y]$, call these the likely antecedents to $[0, n-1; y]$.

In each of these definitions the "likely" objects in sum exhaust most of the probability. In particular, the likely antecedents of $[0, n-1; y]$ exhaust most of the a posteriori probability in $[C, \omega^*]$ of the basic sets $[0, n-1; x]$ when $[0, n-1; y]$ is known. Let us use the word "package" to mean "the aggregate of likely antecedents to a given $[0, n-1; y]$."

8.2. *Marked basic sets.* The nub of Shannon's proof lies in the fact that the packages are so small that it is easy to find 2^{nH} of them which are disjoint. Indeed, suppose one designates, "marks," 2^{nH} of the likely basic sets $[0, n-1; x]$ from $[A, \mu^*]$, doing so at random. Then the probability that a particular $[0, n-1; x]$ be marked in this process is $2^{n(H-H^*)}$. Consider the 2^{nK^*} likely antecedents of some $[0, n-1; y]$. The conditional probability that two or more of these get marked, knowing that one of them is marked, is of the order of

$$2^{nK^*} \cdot 2^{n(H-H^*)} = 2^{n(H-R^*)} \leqq 2^{-n\gamma},$$

by (1) and (2). This probability may be made small by choosing a large n.

8.3. *Distinguishability a posteriori of marked inputs.* Conceptually, we now have this situation: some 2^{nH} basic sets $[0, n-1; x]$ have been specially marked. Given a $[0, n-1; y]$, the received message, and knowing in addition that a marked basic set $[0, n-1; x]$ has been transmitted (has occurred) there is but

a small conditional probability, in the joint universe of $[C, \omega^*]$ and of random markings, that either of the following events has occurred.

(i) The actual $[0, n-1; x]$ which occurred is not a likely antecedent of $[0, n-1; y]$;

(ii) The actual $[0, n-1; x]$ which occurred is a likely antecedent of $[0, n-1; y]$, but there are other marked $[0, n-1; x]$ in the same package.

There is now virtual certainty in the joint universe of $[C, \omega^*]$ and random markings that the actual $[0, n-1; x]$ is a unique marked likely antecedent of $[0, n-1; y]$ when we know $[0, n-1; y]$ a priori, and that a marked $[0, n-1; x]$ is transmitted. That is, by making a marking at random, one is almost certain to have chosen a limited vocabulary of 2^{nH} basic sets $[0, n-1; x]$ which are almost certain to be distinguishable a posteriori, knowing $[0, n-1; y]$.

8.4. *The transducer.* The next step is deceptively simple. One shows easily that, given a marking, a finitary transducer can be described which maps the 2^{nH} likely $[0, n-1; w]$ from $[A_1, \mu]$ on to the marked $[0, n-1; x]$ from $[A, \mu^*]$. When one drives this transducer from $[A_1, \mu]$, the likely output basic sets $[0, n-1; x]$ are just those which were marked. Therefore, when one operates the channel from $[A_1, \mu]$ through this transducer he has essentially only the vocabulary of marked basic sets appearing at the input to the channel. Let us call the resulting joint process of input x to, and output y from, the channel the source $[C, \omega]$. This source itself depends on the marking.

If the probabilities sketched in 8.3 can be relied on for this new situation, it is evident that we have described a transducer, depending on a random marking, which, when $[0, n-1; y]$ is given, permits the correct identification of the $[0, n-1; x]$ which occurred in all but a set of cases of small probability (a posteriori, knowing $[0, n-1; y]$) in the joint universe of random markings and events in $[C, \omega]$. We can assume that for all likely $[0, n-1; x]$ the input $[0, n-1; w]$ which produced it is unique. Then the average, over the joint universe of markings and events in $[C, \omega]$, of the probability that the actual $[0, n-1; w]$ which occurred is not the one determined by this procedure is small. By the Tchebycheff inequality, then, all but a small fraction of the markings will describe transducers which make the probability of misidentifying the actual $[0, n-1; w]$ *simultaneously* small for all but a small fraction of the $[0, n-1; y]$.

8.5. *Critique.* This argument shows that it is somehow easy to describe a transducer which will make the probability of error small. There is, however, a gap in the argument. The probabilities calculated in 8.3 were based on $[C, \omega^*]$. In 8.4 we used these as though they applied to any $[C, \omega]$ which might arise when a marking had been made. If they are both ergodic, and this we are tacitly assuming, ω and ω^* are either identical or else each assigns unit probability to a null set of the other. (This is almost trivial to prove. To my knowledge it was first explicity noted by G. W. Brown.) The probabilities in 8.3 are based on relations which hold only almost everywhere in $[C, \omega^*]$, and therefore, possibly, at most on a null set in $[C, \omega]$. This point is not touched on in [2].

In 10.1 we shall show that finitary channels have a kind of continuity which

permits passage from $[C, \omega^*]$ to $[C, \omega]$, when n is large enough, without serious modification of the probabilities. Shannon's argument is then valid, though incomplete, for finitary channels. Indeed, it is valid for any channel having this kind of continuity, but I have not yet found a satisfying formulation of the property or isolation of the class.

9. The limit theorem.

9.0. *Introduction.* This section is devoted principally to the proof of the theorem quoted in 7.2, which has as a corollary that every ergodic source has the AEP. We recall the definitions of 2.1 and 2.2, and use the following notation.

Given any fixed $x \in A^I$, The symbols $[t, t+n-1; x]$, $[x_t, \cdots, x_{t+n-1}]$ denote that basic set which consists of all $x' \in A^I$ such that $x'_{t+k} = x_{t+k}$, $k = 0, 1, \cdots, n-1$.

Given a source $[A, \mu]$, the symbols $\int f(x)\, d\mu(x)$, $\int f\, d\mu$, denote integration over the space A^I. Integration over a measurable subset $S \subseteqq A^I$ is denoted by one of $\int_S f(x)\, d\mu(x)$, $\int_S f\, d\mu$.

Following [3], we append "(μ)" to a statement which holds almost everywhere with respect to μ, or to a statement involving mean convergence relative to μ.

9.1. *The Theorem.* Given the source $[A, \mu]$ we define the following step functions of $x \in A^I$.

$$
\begin{aligned}
p_n(x) &= \frac{\mu([-n, 0; x])}{\mu([-n, -1; x])}, && n \geqq 1, \\
p_0(x &= \mu([x_0]); \\
g_n(x) &= -\log p_n(x), && n \geqq 0; \\
f_n(x) &= -\frac{1}{n} \log \mu(0, n-1; x]), && n \geqq 1.
\end{aligned} \tag{1}
$$

The function $p_n(x)$ is the conditional probability that the letter which occurs at time $t = 0$ is x_0 when it is known that the letters between time $t = -n$ and $t = -1$ are also those of the infinite sequence x. The definitions of $g_n(x)$ and $f_n(x)$ need no comment. They are related by the important and easily verified formula

$$
f_N(x) = \frac{1}{N} \sum_{k=1}^{N-1} g_k(T^k x). \tag{2}
$$

What is now to be proved is the

THEOREM. *For any source $[A, \mu]$, the sequence $(f_n(x))$ converges in L^1 mean (μ). If $[A, \mu]$ is ergodic, the limit of this sequence is almost everywhere constant and equal to H, the information rate of $[A, \mu]$.*

9.2. PROOF. The proof of this theorem requires the following intermediate results.

(i) The sequence $(p_n(x))$ converges almost everywhere (μ).

(ii) Each $g_n(x) \in L^1(\mu)$, and the sequence $(g_n(x))$ converges in L^1 mean (μ).

These will be established in 9.3 and 9.4, respectively. Granted the second of them, the theorem to be proved follows easily, as we now show.

We have $g_n(x) \in L^1$, and $\lim_n \int | g_n - g | d\mu = 0$ for some function $g \in L^1$. Then the mean ergodic theorem (e.g., [4], equation 2.42) implies that $\sum_{k=0}^{N-1} g(T^k x)/N$ converge in L^1 mean to an invariant function $h(x) = h(Tx)$. When μ is ergodic $h(x) = H$, a constant, almost everywhere.

By (2) of 9.1

$$\int | f_N - h | d\mu \leqq \int \left| \frac{1}{N} \sum_{k=0}^{N-1} [g_k(T^k x) - g(T^k x)] \right| d\mu(x) + \int \left| \frac{1}{N} \sum_{k=0}^{N-1} g(T^k x) - h(x) \right| d\mu(x)$$

$$\leqq \frac{1}{N} \sum_{k=0}^{N-1} \int | g_k(x) - g(x) | d\mu(x) + \int \left| \frac{1}{N} \sum_{k=0}^{N-1} g(T^k x) - h(x) \right| d\mu(x).$$

In the second inequality we use the stationarity of μ to obtain the first term. This term represents the first Cesaro mean of a sequence which by hypothesis has zero as a limit, hence it has also the limit zero. The second term also goes to zero as $N \to \infty$, by the mean ergodic theorem. We conclude then that $f_n \to h$ in L^1 mean, and that $f_n \to H$ in L^1 mean when μ is ergodic. We identify this constant H with the entropy rate of $[A, \mu]$ in 9.4.

9.3. *First Lemma.* We now prove that the sequence $(p_n(x))$ converges almost everywhere (μ). For any given set $D \in F_A$ define, in analogy with 9.1, (1)

$$p_n(x, D) = \frac{\mu([-n, -1; x] \cap D)}{\mu([-n, -1; x])}, \qquad n \geqq 1;$$

this is the conditional probability of D knowing $x_{-n}, x_{-n+1}, \cdots, x_{-1}$. It is a result of Doob [5] that such a sequence of conditional probabilities is a martingale (positive and bounded) and converges almost everywhere.

Given $\alpha \in A$, let D_α denote the basic set of all $x \in A^I$ with $x_0 = \alpha$. Given any $x \in A^I$, the value of $p_n(x)$ is one of the numbers $p_n(x, D_\alpha)$ obtained as α ranges over the finite set A. Therefore

$$| p_n(x) - p_m(x) | \leqq \sum_{\alpha \in A} | p_n(x, D_\alpha) - p_m(x, D_\alpha) |, \tag{3}$$

because the left member is, for each x, one of the summands on the right.

Except for x in a certain null set (μ), each term on the right of (3) converges to zero as m and n go to infinity, by the result of Doob quoted above. By (3), then, the sequence $p_n(x)$ converges almost everywhere (μ), say to $p(x)$.

It follows at once that the sequence $g_n(x) = -\log p_n(x)$ converges almost everywhere to $g(x) = -\log p(x)$, if we admit convergence to $+\infty$.

9.4. *Second Lemma.* We must now show that $g_n(x) \varepsilon L^1$ and that the sequence $g_n(x)$ converges in mean (μ). The integrability of the $g_n(x)$ is simple to establish directly, but will follow automatically from stronger results which are needed later. We need a uniform bound for the contribution of the "unbounded part" of g_n to the value of $\int g_n \, d\mu$. We shall therefore show that

$$\int_{A_{n,L}} g_n \, du \leqq O(L2^{-L}) \tag{4}$$

uniformly in n, where $A_{n,L}$ is the set of x's such that $g_n(x) \geqq L$.

Let $E_{n,K}$ be the set of x's where

$$K \leqq g_n(x) < K + 1. \tag{5}$$

Let B denote a typical basic set $[-n, -1; x]$. Given $\alpha \varepsilon A$, let D_α, as before, be the basic set of all x such that $x_0 = \alpha$. By its definition, $g_n(x)$ is constant over each $B\Lambda D_\alpha$, in fact, it has there the value $-\log[\mu(B\Lambda D_\alpha)/\mu(B)]$. Hence $g_n(x)$ is measurable.

Let a be the number of letters in the alphabet A. There are altogether finitely many, namely a^{n+1}, sets $B\Lambda D_\alpha$ covering A^I. Since $g_n(x) \geqq 0$ everywhere, we have

$$A^I = \bigcup_B \bigcup_{K=0}^{\infty} B\Lambda E_{n,K}.$$

For fixed n, K, let D^K range over those D_α such that $B\Lambda D_{n,K} \neq \phi$. Then the step character of $g_n(x)$ implies that $B\Lambda E_{n,K} = \bigcup_{D^K} B\Lambda D^K$. Therefore

$$\int_{B\Lambda E_{n,K}} g_n \, d\mu = \sum_{D^K} \int_{B\Lambda D^K} g_n \, d\mu \tag{6}$$

and, furthermore, over any $B\Lambda D^K$, (5) holds. Therefore $-\log[\mu(B\Lambda D^K)/\mu(B)] \geqq K$, or $\mu(B\Lambda D^K) \leqq 2^{-K}\mu(B)$. From (6), then,

$$\int_{B\Lambda E_{n,K}} g_n \, d\mu < \sum_{D^K} (K+1)2^{-K}\mu(B) < a(K+1)2^{-K}\mu(B).$$

We have then that

$$\int_{E_{n,K}} g_n \, d\mu = \sum_B \int_{B\Lambda E_{n,K}} g_n \, d\mu < (K+1)2^{-K}, \tag{7}$$

since $\sum \mu(B) = 1$. The right member of (7) is the Kth term of a convergent series and is independent of n. Since $A_{n,L} = \bigcup_{K \geqq L} E_{n,K}$, (4) follows at once.

That $g_n \varepsilon L^1$ follows by summing (7) over all $K \geqq 0$. This summation gives a uniform bound, say β, for $\int g_n \, d\mu$. Define $g_n^L(x) = \inf(g_n(x), L)$, $g^L(x) = \inf(g(x),$

L). Then $\lim_n g_n^L(x) = g^L(x), (\mu)$, and this convergence is dominated by the integrable function L. Hence

$$\lim_n \int | g_n^L - g^L | \, d\mu = 0, \tag{8}$$

and $g^L \varepsilon L^1$. Furthermore

$$\int g^L \, d\mu = \lim_n \int g_n^L \, d\mu \leqq \limsup_n \int g_n \, d\mu \leqq \beta. \tag{9}$$

By (9) and the definition of the left-hand side,

$$\int g \, d\mu = \lim_{L \to \infty} \int g^L \, d\mu \leqq \beta$$

whence $g \varepsilon L^1$. Furthermore

$$\lim_{L \to \infty} \int | g - g^L | \, d\mu = \lim_L \int (g - g^L) \, d\mu = 0. \tag{10}$$

We have now

$$\int | g_n - g | \, d\mu \leqq \int | g_n - g_n^L | \, d\mu + \int | g_n^L - g^L | \, d\mu + \int | g^L - g | \, d\mu.$$

The first term on the right is dominated by

$$\int_{A_{n,L}} g_n \, d\mu$$

where $A_{n,L}$ is the set over which $g_n(x) \geqq L$. By (4) and (8) therefore,

$$0 \leqq \limsup_n \int | g_n - g | \, d\mu \leqq 0(L2^{-L}) + \int | g^L - g | \, d\mu.$$

We let $L \to \infty$ and use (10) to conclude that $\lim_n \int | g_n - g | \, d\mu = 0$.

This establishes the mean convergence of the sequence $(g_n(x))$.

It was shown in 4.3 that the entropy rate of $[A, \mu]$ is $\lim_{n \to \infty} \int g_n \, d\mu$. From what we have just shown, $\lim_{n \to \infty} \int g_n \, d\mu = \int g \, d\mu$. In 9.2, $h(x)$ is defined as the limit of $1/N \sum_{i=0}^{N-1} g(T^i x)$ and we know by the ergodic theorem then that $\int h \, d\mu = \int g \, d\mu$. When μ is ergodic $h(x) = H$, a constant, almost everywhere. Therefore

$$H = \int H \, d\mu = \int h \, d\mu = \int g \, d\mu = \lim \int g_n \, d\mu.$$

This identifies H with the entropy rate of $[A, \mu]$.

10. Finitary devices.

10.0. *Sources.* Shannon's Markov-like sources, which we have here called finitary, are defined by a construction equivalent (in a sense to be made precise later) to one now to be described.

Consider a Markov process with finitely many states, enumerated $1, 2, \cdots, S$. Let p_{ij} be the probability of the transition from state j to state i, and ξ_i the stationary probability of occupancy of state i, so that

$$\xi_i = \sum_{j=1}^{S} p_{ij}\xi_j , \qquad 1 \leqq i \leqq S.$$

Let B be the alphabet whose letters are the symbols $1, 2, \cdots, S$. We may suppose that the Markov process makes a transition at each time $\tau = t + \frac{1}{2}$, $t = 0, \pm 1, \pm 2, \cdots$. We define the stationary information source $[B, \nu]$ by the rule that the letter which occurs at time $t \varepsilon I$ is the name of the state in which the Markov process is at that time. The p_{ij} and the ξ_i are enough to define this source. A source defined in this way will be called a finite Markov source.

Let A be an arbitrary alphabet and let φ be a function from B to A: $\alpha = \varphi(\beta)$, $\beta \varepsilon B$, $\alpha \varepsilon A$. Given $y \varepsilon B^I$, say $y = (\cdots, y_{-1}, y_0, y_1, \cdots)$, we define $x = \Phi(y) \varepsilon A^I$ by $x = (\cdots, x_{-1}, x_0, x_1, \cdots)$, where $x_t = \varphi(y_t)$, $t \varepsilon I$. Let μ be the measure over F_A defined by this construction. In the notation of [3], $\mu = \nu\Phi^{-1}$. The source $[A, \mu]$ we will call a *projection* of $[B, \nu]$. The notion of projection clearly applies even when $[B, \nu]$ is not Markov.

An arbitrary source $[A, \mu]$ will be called finitary if it is a projection of some finite Markov source $[B, \nu]$. Shannon's sources are all of this kind in the sense that, given any source of his, there is a projection of a finite Markov source which produces the same ensemble of text, and conversely.

Consider now an $[A, \mu]$, a projection by φ of $[B, \nu]$. Given $\alpha \varepsilon A$, let $\varphi^{-1}(\alpha)$ denote that subset of B consisting of all β such that $\varphi(\beta) = \alpha$. We will call $[A, \mu]$ *unifilar* if for each $\alpha \varepsilon A$ and each state $i \varepsilon B$ there is at most one transition from i to $\varphi^{-1}(\alpha)$ which has nonzero probability.

The definition of [2], paragraph 7, and certain related results, are tacitly restricted to finitary and unifilar sources. The word "finitary" as we use it in discussing the proof of the fundamental theorem (Section 8) may, however, be interpreted in the wider sense defined above: being a projection of a finite Markov process.

10.1. *Channels.* We now frame a definition of "finitary channel" consonant with that just given for finitary sources. A finitary channel is specified by:

(i) An input alphabet A.

(ii) An output alphabet B.

(iii) A finite set $D = (1, 2, \cdots, K)$ of states. We treat D as an alphabet.

(iv) A set of Markov transition matrices, $\| q_{ij}(\alpha) \|$, one matrix for each $\alpha \varepsilon A$. Each element $q_{ij}(\alpha)$ represents a conditional probability of transition from state $j \varepsilon D$ to state $i \varepsilon D$ knowing that the input letter is α. We have $\Sigma_i\, q_{ij}(\alpha) = 1$ for each j and α.

(v) A function ψ from D to B. The output letter from the channel is $\psi(i)$ whenever the transition is to the state $i \varepsilon D$.

Consider a stationary source $[A, \mu]$ driving the channel so specified. Let λ_j be the stationary probability of finding the channel in state $j \varepsilon D$, (if such a probability exists). Then the probability that the letter α be presented to the channel and that the channel make a transition to state $i \varepsilon D$ is

$$\sum_j \mu([\alpha]) q_{ij}(\alpha) \lambda_j . \tag{1}$$

Stationarity of the system requires now that the sum of these numbers over all $\alpha \varepsilon A$ be λ_i. That is, the vector $(\lambda_i , i \varepsilon D)$ must be invariant under left multiplication by the Markov matrix

$$Q = \| \sum_{\alpha \varepsilon A} \mu([\alpha])\ q_{ij}(\alpha) \| .$$

At least one such invariant probability vector exists. If, for example, each matrix $\| q_{ij}(\alpha) \|$ has a unique such invariant vector, then in general the λ_i will also be unique and they will be continuous functions of the letter frequencies of the source.

Given the λ_i above, the joint probability that letters $\alpha_1 , \cdots , \alpha_n$ be presented to the channel and that the corresponding sequence of states of the channel be $i_1 , i_2 , \cdots , i_n$ is similar to the expression (1):

$$\mu([\alpha_1 , \cdots , \alpha_n]) \sum_{j \varepsilon D} q_{i_n i_{n-1}}(\alpha_n) \cdots q_{i_2 i_1}(\alpha_2) q_{i_1 j}(\alpha_1) \lambda_j . \tag{2}$$

The joint probability of input letters $[\alpha_1 , \cdots , \alpha_n]$ and output letters $[\beta_1 , \cdots , \beta_n]$ is found by summing (2) for

(1) all i_1 in $\psi^{-1}(\beta_1)$,

(2) all i_2 in $\psi^{-1}(\beta_2)$,

.

.

.

(n) all i_n in $\psi^{-1}(\beta_n)$.

The conditional probability of $[\beta_1 , \cdots , \beta_n]$ knowing $[\alpha_1 , \cdots , \alpha_n]$ is then

$$\sum_n \cdots \sum_1 \sum_{j \varepsilon D} q_{i_n i_{n-1}}(\alpha_n) \cdots q_{i_1 j}(\alpha_1) \lambda_j , \tag{3}$$

where $\sum_k$ denotes the summation of i_k over $\psi^{-1}(\beta_k)$. The expression (3) depends on $[\alpha_1 , \cdots , \alpha_n]$ and $[\beta_1 , \cdots , \beta_n]$, and not otherwise upon past history. It is independent of the source except for the continuous dependence of the λ_j upon the letter frequencies. This continuity is sufficient for the proof in Section 8, since it is easy there to guarantee that the source $[A, \mu^*]$ and the source which results from putting $[A_1 , \mu]$ through the transducer there defined have virtually the same letter frequencies.

10.2. *A discrepancy.* The purist will observe that in 2.24 we defined a channel as a set of conditional probability measures ν_θ over outputs, where θ represents the input sequence. The construction in 10.1 is not obviously of this kind, since the measures ν_θ there obtained might well depend not only on θ but also on the particular source-ensemble in mind at the moment. We will not clarify the point here. Some pedagogic license was used in 2.24, and it is simpler to enlarge the notion of channel beyond that defined there than to try to reconcile the two definitions.

11. A useful theorem. Let Ω be an abstract countable set of elements ω. Let A be a finite set, an alphabet. Let μ be a probability measure over a Borel field containing all sets $S \otimes W$, where $S \varepsilon F_A$ and $W \subseteq \Omega$. Define the measures μ_ω over F_A by $\mu_\omega(S) = \mu(S \otimes \omega)/\mu(A^I \otimes \omega)$. This definition is valid for almost every ω. Define $\bar{\mu}$ over F_A by $\bar{\mu}(S) = \mu(S \otimes \Omega)$. Suppose that the source $[A, \bar{\mu}]$ is ergodic and has rate H. Suppose that

$$\int [-\log \mu(A^I \otimes \omega)]\, d\mu(x \otimes \omega) < \infty. \tag{1}$$

Then the functions $f_n(x, \omega) = -(\log \mu([0, n-1; x] \otimes \omega))/n$ converge in L^1 mean to H relative to μ. Considering ω as a parameter, for almost every ω the functions $f_n(x, \omega)$ converge in L^1 mean to H relative to μ_ω.

Proof. Since $\mu([0, n-1; x] \otimes \omega) \leqq \bar{\mu}([0, n-1; x])$ we have

$$f_n(x, \omega) \geqq -\frac{1}{n} \log \bar{\mu}([0, n-1; x]) = g_n(x), \tag{2}$$

where the second equality sign defines $g_n(x)$. Fix n and consider the countable list of events $[0, n-1; x] \otimes \omega$. By the composition law (3.2), extended to infinite sums, the entropy of this list of events is the sum of the entropy of the $[0, n-1; x]$ and the conditional entropy of ω knowing $[0, n-1; x]$:

$$H([0, n-1; x] \otimes \omega) = H([0, n-1; x]) + H_{x,n}(\omega). \tag{3}$$

Now the convexity law (3.3) implies that the average conditional entropy $H_{x,n}(\omega)$ is always less than the unconditional entropy of ω, which latter is the integral asserted to be finite in (1). Hence there is a finite K such that for all n

$$H_{x,n}(\omega) \leqq K. \tag{4}$$

From (2), (3), (4) and the definitions of $\bar{\mu}$ and the entropies,

$$\int |f_n(x, \omega) - g_n(x)|\, d\mu(x \otimes \omega) \int f_n(x, \omega)\, d\mu(x \otimes \omega) - \int g_n(x)\, d\bar{\mu}(x)$$

$$= \frac{1}{n} H([0, n-1; x] \otimes \omega - \frac{1}{n} H([0, n-1; x]) \leqq \frac{1}{n} K.$$

Therefore

$$\int | f_n - H | d\mu \leqq \int | f_n - g_n | d\mu + \int | g_n - H | d\bar{\mu} \leqq \frac{1}{n} K + \int | g_n - H | d\bar{\mu}.$$

Since by hypothesis and (9.1) g_n tends to H in L^1 mean $(\bar{\mu})$, the first conclusion of the theorem follows.

For the second conclusion of the theorem, we note that

$$\int | f_n(x, \omega) - H | d\mu(x \otimes \omega) = \sum_{\omega} \mu(A^I \otimes \omega) \int | f_n(x, \omega) - H | d\mu_\omega(x).$$

Since the left-hand side has limit zero, every term on the right for which $\mu(A^I \otimes \omega) \neq 0$ must have limit zero.

As an application of this theorem let $[B, \nu]$ be a stationary source. Let $[A, \bar{\mu}]$ be a projection by φ of $[B, \nu]$. Let Ω coincide with the alphabet B and for $S \in F_A$ define μ by $\mu(S \otimes \omega) = \nu(\Phi^{-1}(S) \cap D_\omega)$ where D_ω is the set of $y \in B^I$ such that $y_{-1} = \omega$. The theorem then implies (if $[A, \bar{\mu}]$ is ergodic) that the rate of $[A, \mu]$ may be calculated by considering conditional probabilities knowing that the letter ω occurred at time -1. When $[B, \nu]$ is finite and Markov, this often leads to simplified calculations.

As another application, consider a fixed countable partition of A^I into sets $S_\omega \in F_A$. Given an ergodic source $[A, \bar{\mu}]$, define $\mu(S \otimes \omega)$ by $\mu(S \otimes \omega) = \bar{\mu}(S \cap S_\omega)$. The theorem then implies that the entropy rate of $[A, \bar{\mu}]$ can be calculated using only partitions which refine the given one.

REFERENCES

[1] NORBERT WIENER, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, John Wiley and Sons, 1949.

[2] C. E. SHANNON, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27, pp. 379–423, pp. 623–656.

[3] P. R. HALMOS, *Measure Theory*, D. Van Nostrand, 1950.

[4] NORBERT WIENER, "The ergodic theorem," *Duke Mathematical Journal*, Vol. 5, pp. 1–18.

[5] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, 1953.

ON A HEURISTIC METHOD OF TEST CONSTRUCTION AND ITS USE IN MULTIVARIATE ANALYSIS

By S. N. Roy

University of North Carolina

1. Summary. In this paper two closely related heuristic principles of test construction (to be explained in Section 3), called Type I and Type II methods, of which Type II is identified with the usual likelihood ratio method, are noticed as underlying most of the classical tests of hypotheses, simple or composite, on means of univariate normal populations, and on total or partial correlations or regressions in the case of multinormal variates. In these situations the two methods are found to lead to identical tests having properties which happen to be very good in certain cases and moderately good in others. For certain types of composite hypotheses an extension is then made of the Type I method which is applied to construct tests of three different classes of hypotheses on multinormal populations (so as to cover, between them, a very large area of multivariate analysis), yielding in each case a test in general different from the corresponding and current likelihood ratio test. In each case, however, the two tests happen to come out identical for some degenerate "degrees of freedom." In contrast to the likelihood ratio test it is found that in these cases, for general "degrees of freedom," the corresponding Type I test is much easier to use on small samples, because of the relatively greater simplicity of the corresponding small sample distribution problem under the null hypothesis. In each case a lower bound of the power function of the Type I test is also given (against all relevant alternatives), anything like which, so far as the author is aware, would be far more difficult to obtain for the Type II tests in these situations. In this paper the *general* approach to the two methods is entirely of a heuristic nature except that, under fairly wide conditions, a lower bound to the power functions for each of the two types of tests is indicated to be readily available, which, however, is much too crude or wide a bound in *general*.

2. Notation and preliminaries. As far as possible observations and sample quantities will be noted by Roman letters and population parameters by Greek letters; scalars by small letters, matrices by capital letters, column vectors by small letters underscored, and row vectors by priming them; the determinant of a square matrix M by $|M|$; "positive definite" by p.d.; "positive semidefinite" by p.s.d.; "except for a set of points of probability measure zero" or "almost everywhere" by a.e. $A(:p \times q)$ will indicate that the matrix A is $p \times q$, and $I(p)$ will stand for a $p \times p$ unit matrix. $x:N(\xi, \sigma^2)$, $\underline{x}(:p \times 1):N(\underline{\xi}(:p \times 1), \Sigma(:p \times p))$ and $X(:p \times n):N(\underline{\xi}(:p \times 1), \Sigma(:p \times p))$ will indicate respectively that the scalar x is normally distributed about a mean ξ with a variance σ^2, the

Received 12/20/52.

column vector $\underset{\sim}{x}$ is multi-normally distributed about a mean vector $\underset{\sim}{\xi}$ with a (p.d.) covariance matrix Σ, and the n column vectors of the $p \times n$ matrix X are independently and multinormally distributed, each column about a mean vector $\underset{\sim}{\xi}$ with a (p.d.) covariance matrix Σ. Exceptions to this notation will be clearly indicated at the proper places. For the sake of clarity, it may be noted that the X above has the probability density

$$[1 \mid (2\pi)^{-\frac{1}{2}pn} \mid \Sigma \mid^{\frac{1}{2}n}] \exp [-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(X - \xi)(X' - \xi')],$$

where ξ ($:p \times n$) is a $p \times n$ matrix each column of which is the column vector $\underset{\sim}{\xi}$. Furthermore, $d\underset{\sim}{x}$ will stand for $\prod_{i=1}^{p} dx_i$ and dX for $\prod_{\lambda=1}^{n} \prod_{i=1}^{p} dx_{i\lambda}$.

Throughout this paper all *general* discussions will be made in terms of the denumerable case, because I feel that perhaps the ideas are made clearest that way. The extension to the nondenumerable case might in *general* lead to measure theoretic difficulties, but such difficulties do not arise in the applications (most of them being nondenumerable cases) treated here.

The most powerful critical region of size, say $\beta_i (< 1)$, which under fairly general conditions will exist and which under slightly less general conditions will also be unique), of a simple hypothesis H_0 against a simple alternative H_i (such that $H_0, H_i \varepsilon \Omega$, where $i = 1, 2, \cdots$, and where Ω stands for a domain of possibilities) will be denoted by $\omega(H_0, H_i, \beta_i)$, its complement, the acceptance region, by $\omega(H_0, H_i, \beta_i)$, to indicate that in *general* both will depend upon H_0, H_i and β_i. The union of regions $\omega(H_0, H_i, \beta_i)$ over different H_i, β_i or $i(i = 1, 2, \cdots)$ will be denoted by $U_{H_i}\omega(H_0, H_i, \beta_i)$ or simply by $U_i\omega(H_0, H_i, \beta_i)$, and the intersection of regions $\bar{\omega}(H_0, H_i, \beta_i)$ over different $i(i = 1, 2, \cdots)$ by $\bigcap_{H_i}\bar{\omega}(H_0, H_i, \beta_i)$ or simply by $\bigcap_i \bar{\omega}(H_0, H_i, \beta_i)$. $P(H_0, H_i, \beta_i)$ will stand for the power of the most powerful test at level β_i or H_0 against H_i, and will in *general* depend upon all the three elements. It can be easily proved and has been published in an earlier paper [14] that $P(H_0, H_i, \beta_i) > \beta_i$. For convenience a sketch is given here. Assume, for simplicity of discussion but without any essential loss of generality, that we have a set of n continuous stochastic variates, $\underset{\sim}{x}(:n \times 1)$ or simply $\underset{\sim}{x}$, with respective probability densities $\phi_{H_0}(\underset{\sim}{x})$ and $\phi_{H_i}(\underset{\sim}{x})$ (or simply ϕ_{H_0} and ϕ_{H_i}) under the hypotheses H_0 and H_i. Then it is well known that $\omega(H_0, H_i, \beta_i)$ and $\bar{\omega}(H_0, H_i, \beta_i)$ are given respectively by

$$\begin{aligned} &\omega(H_0, H_i, \beta_i) : \phi_{H_i} \geqq \lambda\phi_{H_0}, \\ &\bar{\omega}(H_0, H_i, \beta_i) : \phi_{H_i} < \lambda\phi_{H_0}, \end{aligned} \tag{2.1}$$

where λ is determined by

$$P(\underset{\sim}{x} \varepsilon \omega(H_0, H_i, \beta_i) \mid H_0) : \beta_i.$$

Assume here that ϕ is such that ω defined by (2.1) is unique. Integrating the first inequality of (2.1) over $\omega(H_0, H_i, \beta_i)$ and the second one over $\bar{\omega}(H_0, H_i, \beta_i)$ we have respectively $P(H_0, H_i, \beta_i) \geqq \lambda\beta_i$ and $1 - P(H_0, H_i, \beta_i) < \lambda(1 - \beta_i)$, from which, after a slight reduction, we have

$$P(H_0, H_i, \beta_i) > \beta_i. \tag{2.2}$$

Note that in general λ will be of the form $\lambda(H_0, H_i, \beta_i)$, depending on all the elements. Incidentally, *any* critical region of size β for H_0, whose power with respect to an alternative H is greater than or equal to β, will be called an *unbiased critical region for H_0 against H*.

The likelihood ratio critical region at a level, say α, of H_0 against the whole class $H_i \varepsilon \Omega$, provided that it exists, will be denoted by $\hat{\omega}(H_0, \alpha)$. As is well known it is given by

$$\hat{\omega}(H_0, \alpha) : \phi(\underset{\sim}{x}) \geqq \mu(H_0, \alpha)\phi_{H_0}(\underset{\sim}{x}), \tag{2.3}$$

where, for a given $\underset{\sim}{x}$, $\phi(\underset{\sim}{x})$ stands for the largest $\phi_{H_i}(\underset{\sim}{x})$ (provided that it exists) with respect to variation of H_i over Ω, and where $\mu(H_0, \alpha)$ is given by

$$P(\underset{\sim}{x} \varepsilon \hat{\omega}(H_0, \alpha) \mid H_0) = \alpha. \tag{2.4}$$

Notice that $\phi(\underset{\sim}{x})$ is a function of $\underset{\sim}{x}$ only, being independent of H_i, but may depend on the *total* domain Ω. The power of this test, against any alternative H_i will be denoted by $P(H_0, H_i, \alpha)$.

Assume now that H_0 is a composite hypothesis and $H_i (i = 1, 2, \cdots)$ a composite alternative. In earlier papers [8], [13], [14] the author gave a set of sufficient conditions on ϕ_{H_0} for the availability of similar regions for H_0, and a set of (further) restrictions on ϕ_{H_i} and ϕ_{H_0} for the availability, among these similar regions, of one which is most powerful for H_0 against H_i in the following sense. Suppose H_0 and H_i are composite hypotheses, each characterized by some specified and some unspecified elements, so that, if the unspecified elements were specified, both H_0 and H_i would be simple hypotheses. Now suppose that, among the similar regions for H_0, there is one whose location in the sample space depends on the specified elements of H_0 and possibly on those of H_i, but not on the unspecified elements of H_0 or H_i, but which is nevertheless the most powerful critical region for *any* simple hypothesis within H_0 (obtained by specifying the unspecified elements) against *any* simple alternative within H_i (obtained by specifying the unspecified elements). But this "most powerful" is "most powerful among similar regions." If we drop the restriction of similarity and set up in a straightforward manner the most powerful critical region for the simple hypothesis in question against the simple alternative in question, then we may get a (*nonsimilar*) region having a larger power than that of the most powerful *similar* critical region just referred to. Such a most powerful similar critical region may be conveniently called a *bisimilar* region for H_0 against H_i. The likelihood ratio critical region for composite H_0 against all composite $H_i \varepsilon \Omega$ (which we know how to construct, provided that it exists), can be shown [13], [14] to be a similar region for H_0, under the restrictions just referred to. In this situation the same notation will be used as introduced in the previous paragraph for the case of a simple hypothesis against simple alternatives, and the result (2.2) will also hold, it being noted that, while the regions will be independent of the

unspecified elements in H_0 and H_i, $P(H_0, H_i, \beta_i)$ and $P(H_0, H_i, \alpha)$ however, might depend on the unspecified elements of H_i though not on those of H_0.

3. Type I and Type II tests.

3.1. *Definitions and some remarks.* Consider, for simplicity of discussion but without any essential loss of generality (for the definitions could be immediately carried over into the case of composite hypothesis and alternative) a simple hypothesis H_0 against a simple alternative H_i such that H_0, $H_i (i = 1, 2, \cdots) \in \Omega$.

(*i*) Put $\beta_i = \beta (i = 1, 2, \cdots)$ and set up as the rejection and acceptance regions for H_0 $U_i \omega(H_0, H_i, \beta)$ and its complement $\bigcap_i \bar{\omega}(H_0, H_i, \beta)$ to be called, respectively, U_i and $\bigcap_i$. *This is defined to be a Type I test for H_0 against the whole class $H_i \varepsilon \Omega$*, the level of significance α being given by

$$P(\underset{\sim}{x} \varepsilon U_i \omega(H_0, H_i, \beta) \mid H_0) = \alpha(H_0, \beta), (>\beta). \tag{3.1.1}$$

Let us for the moment assume nontriviality, that is, that given $\alpha < 1$, we can find $\beta = \beta(H_0, \alpha) > 0$, for which (3.1.1) will hold.

(ii) Put, in Section 2, $\lambda(H_0, H_i, \beta_i) = \mu$ (a preassigned constant) for all $i = 1, 2, \cdots$, and rewrite $\omega(H_0, H_i, \beta_i)$ and $\bar{\omega}(H_0, H_i, \beta_i)$ as $\omega^*(H_0, H_i, \mu)$ and $\bar{\omega}^*(H_0, H_i, \mu)$.

Now set up, as the rejection and acceptance regions for H_0, $U_i \omega^*(H_0, H_i, \mu)$ and its complement $\bigcap_i \bar{\omega}^*(H_0, H_i, \mu)$, to be called, respectively, U_i^* and $\bigcap_i^*$, where the β_i's $(i = 1, 2, \cdots)$ are subject to $\lambda(H_0, H_i, \beta_i) = \mu$ (a preassigned constant). *This is defined to be a Type II test for H_0 against the whole class $H_i \varepsilon \Omega$* the level of significance α^* being given by

$$P(\underset{\sim}{x} \varepsilon U_i^*(H_0, H_i, \mu) \mid H_0) = \alpha^*(H_0, \mu). \tag{3.1.2}$$

Here again let us, for the moment, assume nontriviality, that is, that given $\alpha^*(<1)$, we can find a μ such that $\beta(H_0, H_i, \mu) = \beta_i(>0)$ and that (3.1.2) will hold. This can be easily recognized as the likelihood ratio test by the following consideration. Notice that $\omega^*(H_0, H_i, \mu)$ (with a preassigned μ) is given by

$$\omega^*(H_0, H_i, \mu) : \phi_{H_i}(\underset{\sim}{x}) \geqq \mu \phi_{H_0}(\underset{\sim}{x}). \tag{3.1.3}$$

Any $\underset{\sim}{x}$ would belong to $U_i \omega^*(H_0, H_i, \mu)$ if for that $\underset{\sim}{x}$, there were at least one $H_i \varepsilon \Omega$ for which (3.1.3) holds. It is easy to see that this would be accomplished if for that $\underset{\sim}{x}$ the largest $\phi_{H_i}(\underset{\sim}{x})$ (under variation of H_i over Ω) were $\geqq \mu \phi_{H_0}(\underset{\sim}{x})$. Hence it is obvious that

$$\begin{aligned} &U_i \omega^*(H_0, H_i, \mu) : \phi(\underset{\sim}{x}) \geqq \mu \phi_{H_0}(\underset{\sim}{x}) \\ &\bigcap_i \bar{\omega}^*(H_0, H_i, \mu) : \phi(\underset{\sim}{x}) < \mu \phi_{H_0}(\underset{\sim}{x}). \end{aligned} \tag{3.1.4}$$

3.2. *An obvious property of the two types of test.* Notice that U_i includes all $\omega(H_0, H_i, B)$ and U_i^* all $\omega^*(H_0, H_i, \mu)$. Now putting

$$P(\underset{\sim}{x} \varepsilon U_i \mid H_i) \equiv P(U_i, H_i, \alpha) \text{ and } P(\underset{\sim}{x} \varepsilon U_i^* \mid H_i) \equiv P(U_i^*, H_i, \alpha)$$

we shall have from Sections (2) and (3) for the two types of tests

$$(3.2.1)\quad \beta(H_0, \alpha) \equiv \beta < P(H_0, H_i, \beta) \leqq P(U_i, H_i, \alpha) \leqq P(H_0, H_i, \alpha) \leqq 1$$
$$P(H_0, H_i, \alpha) > \alpha$$

$$(3.2.2)\quad \beta^*(H_0, H_i, \alpha) \equiv \beta_i^* < P^*(H_0, H_i, \mu) \leqq P(U_i^*, H_i, \alpha) \leqq P(H_0, H_i, \alpha) \leqq 1$$
$$P(H_0, H_i, \alpha) > \alpha.$$

(3.2.1) and (3.2.2) give respectively, for all $H_i \varepsilon \Omega$, the lower bounds $P(H_0, H_i, \beta)$ and $P^*(H_0, H_i, \mu)$ for $P(U_i, H_i, \alpha)$ and $P(U_i^*, H_i, \mu)$, which, however, in general, would be far from close except sometimes for large "deviations" from H_0. With more knowledge of the forms of ϕ_{H_0} and ϕ_{H_i} it is often possible to get far closer lower bounds; even the actual powers are often computable without much difficulty (and turn out to be pretty high) as for example in most of the classical tests on normal populations.

It is easy to see that the results of (3.1) and (3.2) could be easily generalized to cover the case of composite H_0 against composite $H_i \varepsilon \Omega$ provided that we have similar regions for H_0 and a bisimilar region for H_0 against H_i. This, therefore, need not be separately treated.

3.3. *Display of two classical tests as Type I tests.* (i) Almost all classical tests on univariate and multivariate normal populations (ii) most classical tests on other types of populations and (iii) many tests on multivariate normal populations proposed in recent years are known to be derivable (and indeed many of them have, in fact, been derived) from the "likelihood ratio" principle, so that they belong to Type II. The author finds that all the customary tests in category (i), for example, the test of significance of (1) a mean, (2) a mean difference, (3) total or partial or multiple correlation, and (4) regressions, (5) the F-test in analysis of variance, (6) the test of the hypothesis of equality of standard deviations for two univariate normal populations, (7) the test based on Hotelling's T, all belong to Type I as well. Those classical tests in category (ii) that the author has examined so far also all belong to Type I. Coming to those situations that are sought to be handled by tests proposed under category (iii), the author finds that the likelihood ratio tests offered so far, while they automatically belong to Type II, do not belong to Type I. On the other hand, if, in these situations, one carries out (see Section 5) the spirit and method of discriminant analysis, one gets tests (see Section 6) which belong to Type I in a sense slightly more general than we have indicated so far.

In this section we consider, for illustration, two well known classical tests and show that they belong to Type I.

(i) For $N(\xi_1, \sigma^2)$ and $N(\xi_2, \sigma^2)$ the classical test of $H(\xi_1 = \xi_2) \equiv H_0$ against $H(\xi_1 \neq \xi_2) \equiv H$ at a level α is based on a critical region given by

$$(3.3.1)\quad t \geqq t_0 \quad \text{or} \quad \leqq -t_0,$$

where

$$t \equiv (n_1 + n_2 - 2)^{\frac{1}{2}}\{n_1 n_2/[n_1 + n_2]\}^{\frac{1}{2}}(\bar{x}_1 - \bar{x}_2)/\{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\}^{\frac{1}{2}},$$

and t_0 is given by $P(t \geqq t_0 \mid H_0) = \alpha/2$ and where $(\bar{x}_1, \bar{x}_2)$, (s_1, s_2) stand for the means and standard deviations of two random samples of sizes n_1 and n_2 drawn from $N(\xi_1, \sigma^2)$ and $N(\xi_2, \sigma^2)$, respectively. This is well known as a likelihood ratio test but it is easily checked as Type I as well, in the following way. It is well known that $t \geqq t_0$ is a one-sided uniformly most powerful (bisimilar) region of size $\alpha/2$ for the composite H_0 against the composite $H(\xi_1 > \xi_2) \equiv H_1$ and so also is $t \leqq -t_0$ for H_0 against $H(\xi_1 < \xi_2) \equiv H_2$; taking the union we have (3.3.1) of size α.

(ii) Consider the testing of a general linear hypothesis in analysis of variance which, as is well known, can be *formally* reduced to the following. Suppose we have random samples of sizes n_i, means $\bar{x}_i$ and standard deviations s_i, drawn respectively from $N(\xi_i, \sigma^2)(i = 1, \cdots, k)$, and suppose we want to test $H(\xi_1 = \xi_2 = \cdots = \xi_k) \equiv H_0$ against the whole class H of $(\xi_1, \cdots, \xi_k)$ violating H_0. Put $n \equiv \sum_{i=1}^k n_i$; $\bar{x} \equiv \sum_{i=1}^k n_i\bar{x}_i/n$; $\xi \equiv \sum_{i=1}^k n_i\xi_i/n$. Now the classical F-test for H_0, which is well known to be a likelihood ratio or Type II test has at a level α the critical region given by

$$F \geqq F_0 \tag{3.3.2}$$

where $F \equiv [\sum_{i=1}^k n_i(\bar{x}_i - \bar{x})^2/(k - 1)] \div [\sum_{i=1}^k (n_i - 1)s_i^2/(n - k)]$ and where F_0 is given by $P(F \geqq F_0 \mid H_0) = \alpha$.

To recognize this as a Type I test as well we proceed as follows. It was observed in earlier papers [8], [13] that among similar regions for H_0 (which exist) there is a most powerful (bisimilar) region for H_0 against any specific $(\xi_1, \cdots, \xi_k) \equiv \underline{\xi}$ violating H_0, the region of size, say, β being given by

$$t \geqq t_0 \tag{3.3.3}$$

where $t \equiv \sqrt{n - 2}\cot\theta$;

$$\cos\theta \equiv \sum_{i=1}^k n_i(\bar{x}_i - \bar{x})(\xi_i - \xi) \Big/ \left[\sum_{i=1}^k \{n_i(\bar{x}_i - \bar{x})^2 + (n_i - 1)s_i^2\}\right]^{\frac{1}{2}} \cdot \left[\sum_{i=1}^k n_i(\xi_i - \xi)^2\right]^{\frac{1}{2}}$$

and where t_0 is given by $P(t \geqq t_0 \mid H_0) = \beta$. It was also noticed in those papers that this t has exactly the usual t-distribution with $(n - 2)$ degrees of freedom. Notice that $t_0 \equiv t_0(n, \beta)$ and $\beta \equiv \beta(n, t_0)$. To obtain now the union of regions: $t \geqq t_0$ over different sets of $(\xi_1, \cdots, \xi_k)$ we note that a given set of (observed) $\bar{x}_i$'s and s_i's would belong to the union, if for that set there were at least one t such that $t \geqq t_0$. The union is thus easily checked to be given by: the largest t (by varying over $\xi_1, \cdots, \xi_k$) $\geqq t_0$ (which is fixed). But by (3.3.3) the largest t

would correspond to the largest value of $\cos\theta$, and, given $\bar{x}_i$'s and s_i's, the largest value of $\cos\theta$ (under variation over $\xi_1, \cdots, \xi_k$) is easily seen to be given by:

$$(3.3.4) \quad \cos\theta \equiv \left[\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2\right]^{\frac{1}{2}} \Big/ \left[\sum_{i=1}^{k} \{(n_i - 1)s_i^2 + n_i(\bar{x}_i - \bar{x})^2\}\right]^{\frac{1}{2}},$$

so that the largest t is given by

$$(3.3.5) \qquad t \equiv (n-2)^{\frac{1}{2}}\left[\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2\right]^{\frac{1}{2}} \Big/ \left[\sum_{i=1}^{k} (n_i - 1)s_i^2\right]^{\frac{1}{2}}.$$

Therefore the union of regions: $t \geqq t_0$, is given exactly by (3.3.2), which is the critical region of the F-test. Notice that given the α of the F-test, F_0 is obtained from (3.3.2) in the form $F_0(k-1, n-k; \alpha)$; and next by identifying the union of regions $t \geqq t_0$, with $F \geqq F_0$ we have

$$t_0 \equiv [(k-1)(n-2)F_0/(n-k)]^{\frac{1}{2}} \equiv t_0(k-1, n-k; \alpha);$$

and next from (3.3.3) we have

$$\beta \equiv \beta(n, t_0) \equiv \beta(k-1, n-k; \alpha).$$

3.4. *Some further remarks on the two types of test.* It may be noted (See Sections 2 and 3) that by specializing the β_i's (the sizes of the most powerful critical regions against different alternatives) in *two* special ways we get in a heuristic manner the two types of test. By specializing the β_i's in other ways other heuristic principles could be set up, some of which, in special situations, might be "better" than the Type I or Type II tests. It has already been observed that in many situations Type I and Type II tests would coincide. This does not mean, however, that in those situations, $\beta(H_0, H_i, \alpha)$ of the Type II test would be the β of the Type I test. Given H_0 and the H_i's, it would be possible to find a β for Type I and a μ for Type II such that the same critical region for H_0 against the whole class $H_i \varepsilon \Omega$ could be looked upon as $U_{H_i}\omega(H_0, H_i, \beta)$ in relation to the first type and also as $U_{H_i}\omega^*(H_0, H_i, \mu)$ in relation to the second type.

The following theoretical question or group of questions now under investigation is extremely important. Under what *general* restrictions on the probability law of $\underset{\sim}{x}$ and on H_0 and $H_i \varepsilon \Omega$ would either or both of the tests be nontrivial (in the sense discussed in Section 3) and usable (in the sense of having a distribution problem amenable to tabulation), and unbiased (against all relevant alternatives) and/or admissible and/or reasonably powerful (in the sense of having not too bad a power against all relevant alternatives)? So far as the author is aware, these questions have not yet been adequately discussed in a *general* manner (let alone been answered) even for the likelihood ratio or Type II test (which has so long been extensively used in practice), and no attempt will be made in this paper to discuss these questions. The advantage, however, of having two such heuristic principles (with the possibility of having two different tests in many situations) is that it gives us more elbow room than we would have had with one such principle, in the matter of construction of nontrivial, usable and "pretty good" tests.

4. Extended Type I test (and an obvious property of it). Consider a composite hypothesis H_0 against a set of composite alternatives $H_i \varepsilon \Omega$, $(i = 1, 2, \cdots)$. It often happens, as for example in the three broad situations discussed in Section 5, that, while there are similar regions for H_0, there is among these no most powerful (bisimilar) region for H_0 against any $H_i (i = 1, 2, \cdots)$, but that we have instead the following situation. Suppose we have composite hypotheses $H_{0j} (j = 1, 2, \cdots)$ such that $\bigcap_j H_{0j} \equiv H_0$ and composite alternatives $H_{ij} (i = 1, 2, \cdots ; j = 1, 2, \cdots)$ such that $\bigcap_j H_{ij} \equiv H_i$. Notice that H_{0j} and H_{ij} have more unspecified elements than H_0 and H_i respectively. It may well be that we have (as in the cases discussed in Section 5) not only similar regions for H_{0j} but also, among these, a most powerful (bisimilar) region for H_{0j} against any H_{ij} (one for each i with $j = 1, 2, \cdots$; and $i = 1, 2, \cdots$). Consider critical regions $\omega(H_{0j}, H_{ij}, \beta)$ of size β each. Then by our test procedure, over $\bigcap_j \bigcap_i$ of $\tilde{\omega}(H_{0j}, H_{ij}, \beta)$ (which we call $\bigcap_{ji}$ for simplicity), we are anyway accepting $\bigcap_j H_{0j}$, that is, H_0 and over its complement $U_j U_i \omega(H_{0j}, H_{ij}, \beta)$ we are rejecting at least one H_{0j} and therefore H_0 itself. Suppose we set this up as a heuristic test for H_0 against the whole class $H_i \varepsilon \Omega$. Then the critical region will be $U_j U_i \omega(H_{0j}, H_{ij}, \beta)$ or U_{ji} of size α, given by

$$P(\underset{\sim}{x} \varepsilon U_{ji} \mid H_0) = \alpha \tag{4.1}$$

so that $\alpha \equiv \alpha(H_0, \beta)$ and $\beta \equiv \beta(H_0, \alpha)$. As before, nontriviality will be assumed, and it is easy to check that we shall have for all i and j the following inequality:

$$\beta < P(H_{0j}, H_{ij}, \beta) \leqq P(U_{ji}, H_i, \alpha) \leqq 1. \tag{4.2}$$

It may be noted that while $\omega(H_{0j}, H_{ij}, \beta)$, a bisimilar region of size β for H_{0j} against H_{ij}, is independent of the unspecified elements of H_{0j} and H_{ij} and while the location of U_{ji} must be and its *size might* be (as indeed it is for all the cases considered in Section 5) independent of the unspecified elements of H_{0j} and H_{ij}, $P(H_{0j}, H_{ij}, \beta)$, but *might* involve the unspecified elements of H_{ij} and $P(H_0, H_i, \alpha)$ involve those of H_i. As observed in Section 3, the lower bound to the power of the test, given by (4.2), while it is in general easily available, is, at the same time, much too crude. With more knowledge of the probability law a much closer lower bound can often be found as will be exemplified in later sections.

5. Application to three multivariate problems.

5.1. *Statement of the problems.* Three different types of hypotheses will be discussed here, namely, (i) the hypothesis of equality of covariance matrices of two p-variate normal populations, (ii) the hypothesis of equality of k means for each of p variates for k p-variate normal populations with the same covariance matrix (which is *formally* tied up with the general problem of testing a linear hypothesis), and (iii) the hypothesis that in a $(p_1 + p_2)$-variate normal population the set of, say, the first p_1 variates is uncorrelated with the set of the last p_2 variates. In symbols, using the notation given in Section 2, we can rewrite

these hypotheses as (i) $H(\Sigma_1 = \Sigma_2)(\equiv H_0)$ against *all* $H(\Sigma_1 \neq \Sigma_2)(\equiv H)$, (ii) $H(\xi_1 = \xi_2 \cdots = \xi_k)(\equiv H_0)$ (assuming a common Σ) against *all* $H(\neq H_0)$ (assuming again a common Σ) and (iii) $H(\Sigma_{12} = 0)(\equiv H_0)$ against *all* $H(\Sigma_{12} \neq 0)$ $(\equiv H)$, where the $(p_1 + p_2)$ variates have a covariance matrix Σ of the following structure:

$$\Sigma = \begin{matrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} \begin{matrix} p_1 \\ p_2 \end{matrix} \\ \begin{matrix} p_1 & p_2 \end{matrix} \end{matrix}. \tag{5.1}$$

5.2. *Direct Type I construction not possible.* It is well known that there are infinitely many similar regions for each of the above composite hypotheses but no most powerful (bisimilar) region for $H(\Sigma_1 = \Sigma_2)$ against any *specific* $H(\Sigma_1 \neq \Sigma_2)$ or for $H(\xi_1 = \cdots = \xi_k)(\equiv H_0)$ against any *specific* H violating H_0 or for $H(\Sigma_{12} = 0)$ against any *specific* $H(\Sigma_{12} \neq 0)$, so that direct Type I construction will not work here.

5.3. *Reduction to pseudo-univariate and pseudo-bivariate problems.* At this point suppose that, starting from an $\underset{\sim}{x}(: p \times 1)$ which is $N(\xi, \Sigma)$, we consider a linear compound of $\underset{\sim}{x}$, namely $\mu'\underset{\sim}{x}$ (with an arbitrary constant, that is, nonstochastic $\mu'(:1 \times p)$ of nonzero modulus) which is a scalar well known to be $N(\mu'\xi, \mu'\Sigma\mu)$. Note that $\mu'\xi$ and $\mu'\ \Sigma\mu$ are also scalars. Suppose further that we also start from

$$\underset{\sim}{x} = \begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} \begin{matrix} p_1 \\ p_2 \end{matrix} = N\left(\left(\begin{matrix} \xi_1 \\ \xi_2 \end{matrix}\right) : \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}\right),$$

and consider linear compounds $\mu_1'\underset{\sim}{x}_1$ and $\mu_2'\underset{\sim}{x}_2$ (where $\mu_1(: p_1 \times 1)$ and $\mu_2(: p_2 \times 1)$ are each nonnull); then these two scalars are well known to be distributed as a bivariate normal with a correlation coefficient

$$\rho(\mu_1, \mu_2) \equiv \rho_{12} = \mu_1'\Sigma_{12}\mu_2/[(\mu_1'\Sigma_{11}\mu_1)^{\frac{1}{2}}(\mu_2'\Sigma_{22}\mu_2)^{\frac{1}{2}}]. \tag{5.3.1}$$

Now suppose that, in place of (i), (ii) and (iii) of 5.1, we consider respectively

(iv) $H(\mu'\Sigma_1\mu = \mu'\Sigma_2\mu)(\equiv H_{0\mu})$ against *all* $H(\mu'\Sigma_1\mu \neq \mu'\Sigma_2\mu)(\equiv H_\mu)$, ($\mu$ fixed),
(v) $H(\mu'\xi_1 = \cdots = \mu'\xi_k)(\equiv H_{0\mu})$ against *all* $H_\mu(\neq H_{0\mu})$, (μ fixed) and
(vi) $H(\mu_1'\Sigma_{12}\mu_2 = 0)(\equiv H_{0\mu_1\mu_2})$ against *all* $H(\mu_1'\Sigma_{12}\mu_2 \neq 0)(\equiv H_{\mu_1\mu_2})(\mu_1, \mu_2$ fixed).

We now consider the *totality of all* nonnull μ for (iv) and (v) and *all* nonnull μ_1 and μ_2 for (vi). Notice that (i) $\bigcap_\mu H(\mu'\Sigma_1\mu = \mu'\Sigma_2\mu) = H(\Sigma_1 = \Sigma_2)$, (ii) $\bigcap_\mu H(\mu'\xi_1 = \cdots = \mu'\xi_k) = H(\xi_1 = \cdots = \xi_k)$ and (iii) $\bigcap_{\mu_1\mu_2} H(\mu_1'\Sigma_{12}\mu_2 = 0) = H(\Sigma_{12} = 0)$. We could have worked in terms of any subset of such μ's which leads by intersection to the same H_0, but this we do not do here. It may be noted that by the procedure to be used here, apart from measure-theoretic difficulties which, however, do not arise in these applications, the total set of

μ's or any subset of it (of the kind considered) will *uniquely* define an extended Type I test *associated with the total set or with that particular subset.* Next suppose that, in the alternative, under (iv), (v) and (vi), we substitute "*specific*" for "*all*" and thus have three new situations (vii), (viii) and (ix). It is well known that for each of the situations (vii), (viii) and (ix) we have *one most powerful (bisimilar) region,* so that from these we can construct respective *Type I regions* for the univariate situations (iv) and (v) and the bivariate situation (vi), and from these *Type I tests* we can try to construct the respective *extended Type I tests* for the situations (i), (ii) and (iii). This ties up (in the Section 4) the two p-variate problems (i) and (ii) with the two univariate problems (iv) and (v), and the $(p_1 + p_2)$-variate problem with the bivariate problem (vi).

5.4. *A useful notation and reduction.* For an observation matrix $X(:p \times n)$ with elements $x_{i\lambda}(i = 1, 2, \cdots, p; \lambda = 1, 2, \cdots, n)$, les us put $\bar{x}_i \equiv \sum_{\lambda=1}^{n} x_{i\lambda}/n$, $(i = 2, \cdots, p)$ and $\underset{\sim}{x}'(\equiv \bar{x}_1, \cdots, \bar{x}_p)$. Then the covariance matrix $S(:p \times p)$ will be given by $(n - 1)S = XX' - n\underset{\sim}{x}\underset{\sim}{x}'$. Now suppose that in the situation (i) we have two observation matrices $X_r(:p \times n_r)$, $p \leqq n_r - 1$, two mean vectors $\underset{\sim}{x}_r(:p \times 1)$ and two covariance matrices $S_r(:p \times p)$ such that $(n_r - 1)\ S_r = X_rX_r' - n_r\underset{\sim}{x}_r\underset{\sim}{x}_r'$ $(r = 1, 2)$, so that S_r is always at least p.s.d. In situation (ii) assume that we have k observation matrices $X_r(:p \times n_r)$; mean vectors $\underset{\sim}{x}_r(:p \times 1)$; a grand mean vector $\underset{\sim}{x}(:p \times 1)$ such that $\underset{\sim}{x} = \Sigma n_r\underset{\sim}{x}_r/n$, where $n \equiv \sum_{r=1}^{k} n_r$ and $p \leqq n - k$; a covariance matrix of means $S^*(:p \times p)$ defined by $(k - 1)S^* = XX' - n\underset{\sim}{x}\ \underset{\sim}{x}'$, where $X(:p \times k) \equiv (\sqrt{n_1}\underset{\sim}{x}_1 \cdots \sqrt{n_k}\ \underset{\sim}{x}_k)$; k covariance matrices $S_r(:p \times p)$ such that $(n_r - 1)S_r = X_rX_r' - n_r\underset{\sim}{x}_r\underset{\sim}{x}_r'$; a pooled covariance matrix $S(:p \times p)$ such that $(n - k)S = \sum_{r=1}^{k}(n_r - 1)S_r$ so that both S^* and S must be always at least p.s.d. Finally in situation (iii) suppose that we have an observation matrix X and a mean vector $\underset{\sim}{x}$ given by:

$$X \equiv \begin{matrix} \\ X_1 \\ X_2 \\ \\ n \end{matrix}\begin{matrix} \\ p_1 \\ p_2 \\ \\ \end{matrix} \quad \text{and} \quad \underset{\sim}{x} \equiv \begin{matrix} \\ \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \\ \\ \end{matrix}\begin{matrix} \\ p_1 \\ p_2 \\ \\ \end{matrix},$$

and a covariance matrix $S(:(p_1 + p_2) \times (p_1 + p_2))$ given by:

$$(n - 1)S \equiv (n - 1)\begin{matrix} \\ S_{11} \\ S_{12} \\ \\ p_1 \end{matrix}\begin{matrix} \\ S_{12} \\ S_{22} \\ \\ p_2 \end{matrix} \equiv \begin{matrix} \\ X_1X_1' \\ X_2X_1' \\ \\ \end{matrix}\begin{matrix} \\ X_1X_2' \\ X_2X_2' \\ \\ \end{matrix} - n\begin{matrix} \\ \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \\ \\ \end{matrix}(\underset{\sim}{x}_1'\underset{\sim}{x}_2').$$

Here we observe that S must be always at least p.s.d. and also assume that $p_1 \leqq p_2$ and $p_1 + p_2 \leqq n - 1$.

5.5 *Type I tests for the situations (iv), (v) and (vi).*

(iv) Put $F_{\underset{\sim}{\mu}} = \underset{\sim}{\mu}'S_{12}\underset{\sim}{\mu}/\underset{\sim}{\mu}'S_{22}\underset{\sim}{\mu}$ and notice that, at a level β, for $H(\underset{\sim}{\mu}'\Sigma_{12}\underset{\sim}{\mu} = \underset{\sim}{\mu}'\Sigma_{22}\underset{\sim}{\mu})$ $(\equiv H_{0\underset{\sim}{\mu}})$ against all $H(\underset{\sim}{\mu}'\Sigma_{12}\underset{\sim}{\mu} > \underset{\sim}{\mu}'\Sigma_{22}\underset{\sim}{\mu})$ we have the one-sided uniformly most powerful (bisimilar) region: $F_{\underset{\sim}{\mu}} \geqq F_0$, and for $H(\underset{\sim}{\mu}'\Sigma_{12}\underset{\sim}{\mu} = \underset{\sim}{\mu}'\Sigma_{22}\underset{\sim}{\mu})(\equiv H_{0\underset{\sim}{\mu}})$ against all $H(\underset{\sim}{\mu}'\Sigma_{12}\underset{\sim}{\mu} < \underset{\sim}{\mu}'\Sigma_{22}\underset{\sim}{\mu})$ we have the one-sided uniformly most powerful region:

$F_{\underline{\mu}} \leqq F_0'$, where F_0 and F_0' are given by: $P(F_{\underline{\mu}} \geqq F_0 \mid H_{0\underline{\mu}}) = P(F_{\underline{\mu}} \leqq F_0' \mid H_{0\underline{\mu}}) = \beta$. Notice that this $F_{\underline{\mu}}$ has the ordinary F-distribution with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom. The Type I critical region will now be of size 2β, being given by

$$(5.5.1) \qquad \omega_{\underline{\mu}}: (F_{\underline{\mu}} \geqq F_0) U (F_{\underline{\mu}} \leqq F_0'), \qquad \text{or} \qquad : F_{\underline{\mu}} \geqq F_0 \qquad \text{or} \qquad \leqq F_0'.$$

For $n_1 = n_2$ this will be an unbiased critical region, but, for $n_1 \neq n_2$, this will be biased for certain small deviations and unbiased for all large deviations from the hypothesis. In any case, in this situation it is possible to construct a better (but slightly more difficult test) which will not be discussed here.

(v) For $H(\underline{\mu}'\underline{\xi}_1 = \cdots = \underline{\mu}'\underline{\xi}_k)(\equiv H_{0\underline{\mu}})$ against any *specific* $H_{\underline{\mu}}(\neq H_{0\underline{\mu}})$ there is the most powerful (bisimilar) critical region (discussed in Section 3) (of size, say, γ) which is a one-sided t-region, and by taking the union of these regions (for fixed $\underline{\mu}$ but by variations over $\underline{\xi}_1, \cdots, \underline{\xi}_k$), we have the Type I region given by

$$(5.5.2) \qquad F_{\underline{\mu}} \equiv \underline{\mu}' S^* \underline{\mu} / \underline{\mu}' S \underline{\mu} \geqq F_0,$$

when F_0 is obtained from $P(F_{\underline{\mu}} \geqq F_0 \mid H_{0\underline{\mu}}) = \beta$.
This is also well known to be a Type II or likelihood ratio test having in this situation various good properties (including unbiasedness and admissibility). Notice that this $F_{\underline{\mu}}$ has the ordinary F-distribution with $(k - 1)$ and $(n - k)$ degrees of freedom.

(vi) Put $r_{\underline{\mu}_1\underline{\mu}_2} \equiv \underline{\mu}_1' S_{12} \underline{\mu}_2 / (\underline{\mu}_1' S_{11} \underline{\mu}_1)^{\frac{1}{2}} (\underline{\mu}_2' S_{22} \underline{\mu}_2)^{\frac{1}{2}}$ and notice that, at a level β, for $H(\underline{\mu}_1' \Sigma_{12} \underline{\mu}_2 = 0)$ $(\equiv H_{0\underline{\mu}_1\underline{\mu}_2})$ against all $H(\underline{\mu}_1' \Sigma_{12} \underline{\mu}_2 > 0)$ we have the one-sided uniformly most powerful (bisimilar) region: $r_{\underline{\mu}_1\underline{\mu}_2} \geqq r_0$ and for $H_{0\underline{\mu}_1\underline{\mu}_2}$ against all $H(\underline{\mu}_1' \Sigma_{12} \underline{\mu}_2 < 0)$ we have the one-sided uniformly most powerful (bisimilar) region $r_{\underline{\mu}_1\underline{\mu}_2} \leqq -r_0$, where r_0 is given by:

$$(5.5.3) \qquad P(r_{\underline{\mu}_1\underline{\mu}_2} \geqq r_0 \mid H_{0\underline{\mu}_1\underline{\mu}_2}) = \beta.$$

Notice that $r_{\underline{\mu}_1\underline{\mu}_2}$ has the distribution of the ordinary total correlation coefficient on a sample of size n. The Type I critical region will be of size 2β, being given by

$$(5.5.4) \qquad \omega_{\underline{\mu}_1\underline{\mu}_2}: (r_{\underline{\mu}_1\underline{\mu}_2} \geqq r_0)\ U (r_{\underline{\mu}_1\underline{\mu}_2} \leqq -r_0),$$

that is, $\mid r \mid \geqq r_0 \mid$ or $r^2 \geqq r_0^2$.
This is well known to be also a Type II or likelihood ratio region having in this situation various good properties (including unbiasedness and admissibility).

5.6. *Actual construction of extended Type I tests for the situations (i), (ii) and (iii).*

(i) By the test procedure (5.5.1), over $F_0' < F_{\underline{\mu}} < F_0$ we accept $H(\underline{\mu}'\Sigma_1\underline{\mu} = \underline{\mu}'\Sigma_2\underline{\mu})$ so that over $\bigcap_{\underline{\mu}}[F_0' < F_{\underline{\mu}} = \underline{\mu}'S_1\underline{\mu}/\underline{\mu}'S_2\underline{\mu} < F_0]$ we accept $\bigcap_{\underline{\mu}} H(\underline{\mu}'\Sigma_1\underline{\mu} = \underline{\mu}'\Sigma_2\underline{\mu}) \equiv H(\Sigma_1 = \Sigma_2) \equiv H_0$, and thus over its complement $U_{\underline{\mu}}[F_{\underline{\mu}} \geqq F_0$ or $\leqq F_0']$ we reject H_0. This may thus be set up as the extended Type I test. To obtain $U_{\underline{\mu}}[F_{\underline{\mu}} \geqq F_0$ or $\leqq F_0']$ we note that a particular set of observations, that is, a particular set of (S_1, S_2) would belong to the union if for that (S_1, S_2) there were at least one $\underline{\mu}$ such that $F_{\underline{\mu}} \geqq F_0$ or $\leqq F_0'$. It is thus easy to check that $U_{\underline{\mu}}[F_{\underline{\mu}} \geqq$

F_0 or $\leqq F_0'$] is precisely equivalent to: the largest $F_{\underline{\mu}} \geqq F_0$ and/or the smallest $F_{\underline{\mu}} \leqq F_0'$, the "largest" and the "smallest" being under variation of $\underline{\mu}$ (for a given set of S_1, S_2). Now, given (S_1, S_2), the largest and smallest value of $\underline{\mu}'S_1\underline{\mu}/\underline{\mu}'S_2\underline{\mu}$ are easily seen to be the largest and smallest roots, say θ_p and θ_1, of the p-th degree determinantal equation in θ

$$| S_1 - \theta S_2 | = 0, \tag{5.6.1}$$

all the p roots $\theta_1, \theta_2, \cdots, \theta_p$ being in this situation a.e. positive, since S_1 and S_2 are by the definitions and assumptions of subsection 5.4 of Section 5, a.e., p.d. (each of rank p). Starting out from the Type I test (5.5.1) for $H_{0\underline{\mu}}$ we have for $H(\Sigma_1 = \Sigma_2)$ the extended Type I critical region

$$\theta_p \geqq F_0 \text{ and/or } \theta_1 \leqq F_0'. \tag{5.6.2}$$

To determine the size of this critical region, or more properly, given the size α, to find F_0 and F_0', we have to have the joint distribution of $(\theta_1, \theta_2, \cdots, \theta_p)$ on the null hypothesis $H(\Sigma_1 = \Sigma_2)$ which was obtained in 1939 by a number of workers [3], [4], [7], [10] and which was found to be independent of the common value of $\Sigma_1 = \Sigma_2$ and also of $\underline{\xi}_1$ and $\underline{\xi}_2$, that is, of all nuisance parameters. Starting from the joint distribution of $(\theta_1, \cdots, \theta_p)$ on the null hypothesis, we can obtain, by a technique given in earlier papers [9], [12], the joint distribution of (θ_1, θ_p), from which F_0 and F_0' will be available, in terms of α, by using

$$P(\theta_p \geqq F_0 \mid \Sigma_1 = \Sigma_2) = P(\theta_1 \leqq F_0' \mid \Sigma_1 = \Sigma_2), \text{ and } P(\theta_p \geqq F_0 \text{ and/or } \theta_1 \leqq F_0' \mid \Sigma_1 = \Sigma_2) = \alpha. \tag{5.6.3}$$

(ii) By the test procedure (5.5.2), over $F_{\underline{\mu}} \equiv \underline{\mu}'S^*\underline{\mu}/\underline{\mu}'S\underline{\mu} < F_0$ accept $H(\underline{\mu}'\underline{\xi}_1 = \cdots = \underline{\mu}'\underline{\xi}_k)$, so that over $\bigcap_{\underline{\mu}}[F_{\underline{\mu}} < F_0]$ we accept $\bigcap_{\underline{\mu}} H(\underline{\mu}'\underline{\xi}_1 = \cdots \underline{\mu}'\underline{\xi}_k) \equiv H(\underline{\xi}_1 = \cdots = \underline{\xi}_k) \equiv H_0$, and over its complement $U_{\underline{\mu}}[F_{\underline{\mu}} \geqq F_0]$ we reject H_0. We set it up as the extended Type I test for H_0, and note, as before, that $U_{\underline{\mu}}[F_{\underline{\mu}} \equiv \underline{\mu}'S^*\underline{\mu}/\underline{\mu}'S\underline{\mu} \geqq F_0]$ is precisely equivalent to: the largest $\underline{\mu}'S^*\underline{\mu}/\underline{\mu}'S\underline{\mu} \geqq F_0$, the "largest" being under variation of $\underline{\mu}$ (for a given set of observations, that is, for a given set of S^* and S). As before, given S^* and S, the largest value of $\underline{\mu}'S^*\underline{\mu}/\underline{\mu}'S\underline{\mu}$ is checked to be the largest root θ_q of the p-th degree determinantal equation in θ

$$| S^* - \theta S | = 0. \tag{5.6.4}$$

From the definitions and assumptions of (subsection 5.4) of Section 5, it is easy to check that S is, a.e., p.d. while S^* is, a.e., at least p.s.d. of rank $q \equiv \min(p, k-1)$. It will of course be, a.e., p.d. if $p \leqq k-1$. In any case we can say that, of the p roots of (5.41), $p - q$ will be always zero, while q roots, to be called $\theta_1, \cdots, \theta_q$, will be, a.e., positive, where $q \equiv \min(p, k-1)$, so that $0 < \theta_1 \leqq \cdots \leqq \theta_q < \infty$ (suppose). The extended Type I critical region for H_0 is thus

$$\theta_q \geqq F_0. \tag{5.6.5}$$

To determine the size of the region, or rather, given the size α of (5.6.5), to determine F_0, we observe what was noted in the earlier papers [4], [7], [10], namely, that the joint distribution of $(\theta_1, \cdots, \theta_q)$ (on the null hypothesis) in this case is exactly of the same form as that of $(\theta_1, \cdots, \theta_p)$ of the previous case (on the null hypothesis in that situation) and that therefore the distribution of any root, say the largest, will come through by the same technique as was mentioned for the previous case and will also be independent of all nuisance parameters. We shall thus have F_0 given, in terms of α, by,

$$P(\theta_q \geqq F_0 \mid \xi_1 = \cdots = \xi_k) = \alpha. \tag{5.6.6}$$

For $k = 2$ we shall have $q = 1$, so that there will be just one nontrivial sample root $\theta_q (\equiv \theta$ suppose), and just one nontrivial population root $\Theta_r (\equiv \Theta$ suppose) (which will be zero on the null hypothesis and $\neq 0$, on the nonnull hypothesis). This θ is easily checked to be Hotelling's T^2 and its distribution both on the null and nonnull hypothesis are well known [1], [5] and relatively easy, so that (5.6.5) and (5.6.6) happen to be computationally much simpler in this situation.

(iii) By the test procedure (5.5.4), over $r^2_{\underline{\mu}_1\underline{\mu}_2} \equiv (\underline{\mu}_1' S_{12}\underline{\mu}_2)^2/[(\underline{\mu}_1' S_{11}\underline{\mu}_1)(\underline{\mu}_2' S_{22}\underline{\mu}_2)] \geqq r_0^2$, we reject $H(\underline{\mu}_1' S_{12}\underline{\mu}_2 = 0) \equiv H_{0\underline{\mu}_1\underline{\mu}_2}$ and over its complement accept this hypothesis, so that over $\bigcap_{\underline{\mu}_1\underline{\mu}_2}[r^2_{\underline{\mu}_1\underline{\mu}_2} < r_0^2]$ we accept $\bigcap_{\underline{\mu}_1\underline{\mu}_2} H(\underline{\mu}_1' \Sigma_{12}\underline{\mu}_2 = 0) \equiv H_0$, and over its complement $U_{\underline{\mu}_1\underline{\mu}_2}[r^2_{\underline{\mu}_1\underline{\mu}_2} \geqq r_0^2]$ reject H_0. We set this up as the extended Type I test for H_0 and note that

$$U_{\underline{\mu}_1\underline{\mu}_2}[r^2_{\underline{\mu}_1\underline{\mu}_2} \equiv (\underline{\mu}_1' S_{12}\underline{\mu}_2)^2/(\underline{\mu}_1' S_{11}\underline{\mu}_1)(\underline{\mu}_2' S_{22}\underline{\mu}_2) \geqq r_0^2]$$

is exactly equivalent to: the largest value of

$$(\underline{\mu}_1' S_{12}\underline{\mu}_2)^2/(\underline{\mu}_1' S_{11}\underline{\mu}_1)(\underline{\mu}_2' S_{22}\underline{\mu}_2) \geqq r_0^2,$$

the "largest" being under variation of $\underline{\mu}_1$ and $\underline{\mu}_2$ (for a given set of observations, that is, for a given set of S_{11}, S_{22} and S_{12}). As before, the largest value of this expression is checked to be the largest root θ_{p_1} of the p_1st degree determinantal equation in θ

$$| \theta S_{11} - S_{12} S_{22}^{-1} S_{12}' | = 0. \tag{5.6.7}$$

From the definitions and assumptions of subsection 5.4 of Section 5, it is easy to see that S and, therefore, S_{11} and S_{22} are, a.e., p.d. and S_{12} is, a.e., of rank p_1. Under these conditions it is well known and proved in a number of places [6], [15] that the p_1 roots of (5.6.7) will all, a.e., lie between 0 and 1, satisfying, say, $0 < \theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_{p_1} < 1$. The extended Type I region for H_0 is thus

$$\theta_{p_1} \geqq r_0^2. \tag{5.6.8}$$

To determine the size of this region, or rather, given the size α, to determine r_0^2, we observe that the joint distribution of $(\theta_1, \cdots, \theta_{p_1})$ on the null hypothesis in this case goes over (under a simple transformation from cosine to cotangent) into that of the joint distribution of the roots (on the respective null hypotheses) in the two previous cases and the same technique for finding the distribution of

the largest root also goes through. As before, this distribution will also be independent of all nuisance parameters. We shall thus have r_0^2 given, in terms of α, by,

$$P(\theta_{p_1} \geqq r_0^2 \mid \Sigma_{12} = 0) = \alpha. \tag{5.6.9}$$

6. Lower bounds of the powers of the test regions (5.6.2), (5.6.5) and (5.6.8) for the hypotheses (i), (ii), and (iii).

6.1. *Observations on the actual power functions.*

(i) It is well known that on the nonnull hypothesis the joint distribution of $(\theta_1, \cdots, \theta_p)$ of (5.6.1) (and hence of $(\theta_1$ and $\theta_p)$) also involves as parameters only the p roots $\Theta_1, \cdots, \Theta_p$ of the population determinantal equations in Θ,

$$|\Sigma_1 - \Theta\Sigma_2| = 0. \tag{6.1.1}$$

(Notice that, assuming Σ_1 and Σ_2 to be both p.d., these roots will all be positive and they will all be unity if and only if $\Sigma_1 = \Sigma_2$, that is, on the null hypothesis in this situation.) The exact distribution of $(\theta_1, \cdots, \theta_p)$ or of (θ_1, θ_p) on the nonnull hypothesis will be quite complicated and whatever reduction is already known to be possible [11], will not be discussed here. We shall merely write the power function formally as:

$$P[\theta_p \geqq F_0 \text{ and/or } \theta_1 \leqq F_0' \mid \Sigma_1 \neq \Sigma_2] \equiv P\{\alpha; n_1, n_2, p; \Theta_1, \Theta_2, \cdots, \Theta_p\}, \tag{6.1.2}$$

to indicate on which parameters the power depends.

(ii) To discuss the power function of the region (5.6.5), we use the convenient notation: $\underset{\sim}{\xi} \equiv \sum_{r=1}^{k} n_r\xi_r/n$; $\xi(:p \times k) \equiv (\sqrt{n_1}\xi_1, \cdots, \sqrt{n_k}\xi_k)$; $(k-1)\Sigma^* \equiv \xi\xi' - n\underset{\sim}{\xi}\underset{\sim}{\xi}'$; denote by Σ the (assumed) common p.d. covariance matrix of the k populations. We note that $\Sigma^*(:p \times p)$ is p.s.d. (and might also be p.d.) of rank $r \leqq \min(p, k-1)$, where r is the rank of the matrix,

$$(\sqrt{n_1}(\xi_1 - \underset{\sim}{\xi}), \cdots, \sqrt{n_k}(\xi_k - \underset{\sim}{\xi})).$$

Notice that the rank of this matrix must be $\leqq \min(p, k-1)$. Notice further that Σ^* will be zero if and only if $\xi_1 = \xi_2 = \cdots = \xi_k$, that is, on the null hypothesis in this situation. We next observe, as is well known, that on the nonnull hypothesis the joint distribution of $(\theta_1, \cdots, \theta_q)$ of (5.6.4) (and also of θ_q) will involve as parameters only the $r(\leqq q \equiv \min(p, k-1))$ positive (the $p - r$ others being zero) roots of the p-th degree population determinantal equation in Θ,

$$|\Sigma^* - \Theta\Sigma| = 0. \tag{6.1.3}$$

As in the previous cases, so also here, the exact distribution of $(\theta_1, \cdots, \theta_q)$ or of θ_q on the nonnull hypothesis will be quite complicated *and also* different from that of the previous situation and whatever reduction is already known to be possible will, as before, not be discussed here. We shall again formally write the power function as

$$P[\theta_q \geqq F_0 \mid \text{under violation of } H(\xi_1 = \cdots = \xi_k)] \equiv P\{\alpha; n, k, p; \Theta_1, \cdots, \Theta_r\}, \tag{6.1.4}$$

to indicate the dependence on the relevant parameters. When $k = 2$ we have $q = 1$, $r = 1$, and in this case (6.1.4) will be the power function of Hotelling T-test, which is computationally quite manageable.

(iii) To discuss the power function of the region (5.6.8), we observe what is well known, namely that, on the nonnull hypothesis, the joint distribution of $(\theta_1, \cdots, \theta_{p_1})$ of (5.6.7) involves as parameters only the roots of the p_1-th degree population determinantal equation in Θ,

$$| \Theta\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}' | = 0. \tag{6.1.5}$$

Assuming a p.d. Σ, it is also known that Σ_{11} and Σ_{22} are both p.d., all roots being less than 1 and q roots being positive and $p_1 - q$ being zero, where q is the rank of $\Sigma_{12}(q \leqq p_1 \leqq p_2)$. We write them as: $0 < \Theta_1 \leqq \cdots \leqq \Theta_q < 1$. We shall not further discuss the complicated nonnull distribution of $(\theta_1, \cdots, \theta_{p_1})$ or of θ_{p_1}, but merely write down formally the power function of the critical region (5.6.8) as,

$$P[\theta_{p1} \geqq r_0^2 \mid \Sigma_{12} \neq 0] \equiv P\{\alpha; n, p_1, p_2; \Theta_1, \cdots, \Theta_q\}, \tag{6.1.6}$$

to indicate the dependence on the relevant parameters.

Although the exact nonnull distributions and hence the exact power functions would be quite complicated in all the foregoing cases we could, if we wanted to, obtain lower bounds, by using (4.2) and noting that the nonnull distribution for the univariate situations (iv) and (v) associated with (i) and (ii), and the bivariate situation (vi) associated with (iii), are all known in computationally manageable forms. But it is possible, as is shown in the next two subsections (6.2) and (6.3), to obtain much closer lower bounds to the power functions (6.1.2), (6.6.4) and (6.1.6). This is accomplished as follows.

6.2. *On invariance and independence.* It is well known [15] that

(i) the roots of (5.6.1) are invariant under the transformation

$$S_1(:p \times p) = \mu(:p \times p)V_1(:p \times p)\mu'(:p \times p) \quad \text{and}$$
$$S_2(:p \times p) = \mu(:p \times p)V_2(:p \times p)\mu'(:p \times p),$$

when μ is any constant (i.e., nonstochastic) nonsingular transformation matrix,

(ii) the roots of (5.6.4) are invariant under the transformation:

$$S^*(:p \times p) = \mu(:p \times p)V^*(:p \times p) \times \mu'(:p \times p) \text{ and}$$
$$S(:p \times p) = \mu(:p \times p)V(:p \times p)\mu'(:p \times p),$$

where μ is any constant nonsingular transformation matrix, and finally

(iii) the roots of (5.6.7) are invariant under the transformation:

$$S_{11}(:p_1 \times p_1) = \mu_1(:p_1 \times p_1)V_{11}(:p_1 \times p_1) \times \mu_1'(:p_1 \times p_1),$$
$$S_{22}(:p_2 \times p_2) = \mu_2(:p_2 \times p_2)V_{22}(:p_2 \times p_2) \times \mu_2'(:p_2 \times p_2)$$

and

$$S_{12}(:p_1 \times p_2) = \mu_1(:p_1 \times p_1)V_{12}(:p_1 \times p_2) \times \mu_2'(:p_2 \times p_2),$$

where μ_1 and μ_2 are any two constant nonsingular transformation matrices.

We next notice that

(i) there exists a nonsingular transformation matrix (not necessarily unique),

$$\mu(:p \times p) \equiv (\mu_1, \cdots, \mu_p)p,$$

under which $\mu\Sigma_1\mu' = D_\Theta$ and $\mu\Sigma_2\mu' = I(p)$ (where D_Θ is a $p \times p$ diagonal matrix whose diagonal elements are $\Theta_1, \cdots, \Theta_p$) and which transforms the p original variates into p new variates distributed in a canonical form, so that, for this set of p μ_i's$(i = 1, 2, \cdots, p)$, $(\mu_i'S_1\mu_i/\mu_i'S_2\mu_i)/(\mu_i'\Sigma_1\mu_i/\mu_i'\Sigma_2\mu_i)$, that is, $(\mu_i'S_1\mu_i/\mu_i'S_2\mu_i)\ \Theta_i(i = 1, \cdots, p)$ will be distributed as p independent F's, each with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom,

(ii) there exists a nonsingular matrix (not necessarily unique), $\mu(:p \times p) \equiv (\mu_1, \cdots, \mu_p)p$, under which $\mu\Sigma^*\mu' = D_\Theta$ and $\mu\Sigma\mu' = I(p)$ (where D_Θ is a diagonal matrix, of whose p diagonal elements, $p - r$ are exactly zero, while the rest, r in number, are $\Theta_1, \cdots, \Theta_r > 0$), and, furthermore, that this transforms the p original variates into p new variates distributed in a canonical form, so that for this set of p μ_i's$(i = 1, 2, \cdots, p)$, $(\mu_i'S^*\mu_i/\mu_i'S\mu_i)(i = 1, 2, \cdots, p)$ will be distributed as p independent F's each with $(k - 1)$ and $(n - k)$ degrees of freedom. We note that out of these p F's, $p - r$ are necessarily central F's (i.e., with "deviation parameters" equal to zero) and r F's are noncentral with "deviations parameters", $(\Theta_1, \cdots, \Theta_r)$ and

(iii) there exist nonsingular matrices (none necessarily unique),

$$\mu_1(:p_1 \times p_1) \equiv (\mu_{11} \cdots \mu_{p_11})p_1,$$

$$\mu_2(:p_2 \times p_2) \equiv (\mu_{12} \cdots \mu_{p_22})p_2,$$

under which $\mu_1\Sigma_{11}\mu_1' = I(p_1)$, $\mu_2\Sigma_{22}\mu_2' = I(p_2)$ and

$$\mu_1\Sigma_{12}\mu_2' = (D_{\sqrt{\Theta}}\ 0)p_1$$

(where $D_{\sqrt{\Theta}}$ is a $p_1 \times p_1$ diagonal matrix of whose diagonal elements, $p_1 - q$ are zero and the rest are nonzero, being $\Theta_1, \cdots, \Theta_q$), and which transforms the original $(p_1 + p_2)$ variates into two new sets of p_1 and p_2 variates, jointly distributed in a canonical form with covariance matrix:

$$\begin{matrix} \begin{pmatrix} I(p_1) & (D_{\sqrt{\Theta}}\ 0) \\ \begin{pmatrix} D_{\sqrt{\Theta}} \\ 0 \end{pmatrix} & I(p_2) \end{pmatrix} & \begin{matrix} p_1 \\ p_1 \\ p_2 - p_1 \end{matrix} \\ p_1 \quad p_1 \quad p_2 - p_1 . & \end{matrix}$$

This means that from the sets $\mu_{i1}(i = 1, 2, \cdots, p_1)$ and $\mu_{j2}(j = 1, 2, \cdots, p_2)$ it is possible to pick out linked μ_{i1} and $\mu_{i2}(i = 1, 2, \cdots, p_1)$ such that

$(\underline{\mu}'_{i1}S_{12}\underline{\mu}_{i2})^2/(\underline{\mu}'_{i1}S_{11}\underline{\mu}_{i1})(\underline{\mu}'_{i2}S_{22}\underline{\mu}_{i2})(i = 1, 2, \cdots, p_1)$ are distributed as the squares of p_1 independent correlation coefficients r_i with $(n - 2)$ degrees of freedom each, the distributions involving $\Theta_i \equiv \rho_i^2 (i = 1, 2, \cdots, q \leqq p_1)$ as "deviation parameters". The absolute value of the total correlation coefficient will be indicated by enclosing the correlation in vertical bars. It is the distribution of this, that is, the distribution of multiple correlation when $p = 2$, that will come into the picture. It is possible to go even beyond this and pick out linked $\underline{\mu}_{i1}$ and $\underline{\mu}_{i2} (i = 1, 2, \cdots, p_1 - 1)$, and at the last stage a $\underline{\mu}_{p_1 1}$ linked with a set of $(p_2 - p_1 + 1)$ $\underline{\mu}_{i2}$'s$(i = p_1, p_1 + 1, \cdots, p_2)$, such that there are p_1 independently distributed | correlations |, of which $(p_1 - 1)$ are | total correlations |, and the last one is a multiple correlation between the p_1th variate of the first p_1-set and the $(p_1, p_1 + 1, \cdots, p_2)$ variates of the second p_2-set. The deviation parameters being $\Theta_i (0 < \Theta_1 \leqq \cdots \leqq \Theta_q < 1)$, we could so arrange that the first $p_1 - q$ sample (total) | correlations | had zero deviation parameters to go with, the next $q - 1$ sample (total) | correlations | had respective (and one each) deviation parameters $(\Theta_1, \cdots, \Theta_{q-1})$ to go with and the last sample (multiple) correlation had Θ_q to go with.

6.3. *Actual construction of lower bounds.* Now notice that

(i) in the first problem, the region (5.6.2) includes as well all the F-regions considered under (*i*) of the foregoing subsection (6.2), so that, to the power function P of (6.1.2) we shall have a lower bound given by

$$(6.3.1) \quad P\{\alpha; n_1, n_2, p; \Theta_1, \cdots, \Theta_p\} > 1 - \prod_{i=1}^{p} [1 - P(F \geqq F_0 \text{ or } \leqq F'_0 \mid \Theta_i)]$$

(each with $n_1 - 1$ and $n_2 - 1$ degrees of freedom), which is easily calculable.

(ii) in the second problem, the region (5.6.5) includes as well all the F-regions considered under (ii) of the preceding subsection 6.2, so that, to the power function P of (6.1.4) we shall have a lower bound given by

$$(6.3.2) \quad \begin{aligned} &P\{\alpha; n, k, p; \Theta_1, \cdots, \Theta_r\} \\ &> 1 - [1 - P(\text{central } F \geqq F_0)]^{p-r} \prod_{i=1}^{r} [1 - P(\text{noncentral } F \geqq F_0 \mid \Theta_i)] \end{aligned}$$

(each with $k - 1$ and $n - k$ degrees of freedom), which is easily calculated; and finally

(iii) in the third problem, the region (5.6.8) includes as well all the | correlation | regions considered under (iii) of the foregoing subsection 6.2, so that, to the power function P of (6.1.6) we shall have a lower bound given by

$$(6.3.3) \quad \begin{aligned} &P\{\alpha; n, p_1, p_2; \Theta_1, \cdots, \Theta_q\} \\ &\qquad > 1 - [1 - P(r^2 \geqq r_0^2 \mid \text{null hypothesis})]^{p_1 - q} \\ &\qquad\qquad \times \prod_{i=1}^{q} [1 - P(r^2 \geqq r_0^2 \mid \rho_i^2 = \Theta_i)], \end{aligned}$$

(each with $n - 2$ degrees of freedom), which is easily calculable, being really the power function of the multiple correlation of the first kind [2], when $p = 2$, for which tables are in part available which could easily be extended with modern computing facilities.

The lower bound (6.3.3) could be easily improved, when $p_2 > p_1$, by the following consideration. Going back to the observations made at the end of subsection 6.2 of this section (on independence between two sets of variates), we notice that since the region (5.6.5) includes $p_1 - 1$ | (*total*) correlation | regions and one (*multiple*) correlation region we shall have a lower bound (easily checked to be larger than (6.3.3)) given by

$$P\{\alpha; n, p, p_2; \Theta_1, \cdots, \Theta_q\} > 1 - [1 - P(r^2 \geqq r_0^2 \mid \text{null hypothesis})]^{p_1 - q} \times \prod_{i=1}^{q-1} [1 - P(r^2 \geqq r_0^2 \mid \rho_i^2 \equiv \Theta_i)] \times [1 - P(R^2 \geqq r_0^2 \mid \rho_q^2 \equiv \Theta_q)], \tag{6.3.4}$$

where all factors except the last are on | total correlations | distributed with $(n - 2)$ degrees of freedom, while the last factor is on a multiple correlation distributed with $(n - 2)$ degrees of freedom and $(p_2 - p_1)$.

It may be noted that in (6.3.2) *both* sides of the inequality are "known," that is, computationally accessible when $k = 2$, that is, $q = 1$ and $r = 1$, the left-hand side being just the power function of Hotelling's T, while the right hand is also easily available (in this as in all other cases).

7. Concluding remarks. It is of considerable importance at this stage to ask how "good" the lower bounds indicated in (6.3.1), (6.3.2) and (6.3.3) or (6.3.4) are. A lower bound to the power could be said to be "good" if it were (i) close to the actual power, and/or (ii) if it were itself pretty large, being greater than the level of significance α for reasonably large values of the deviation parameters and possibly getting larger as those parameters increase. For all the three tests condition (ii) has been numerically checked to be true over a fairly wide range of test values of the several parameters involved, and part of that material will be offered in a later paper. With regard to condition (i), *in general*, that is, *for small samples*, not only do we not know the actual power (in which case the search for a lower bound would have been redundant) but at the moment we do not even know an upper bound of the expression: (actual power − given lower bound to it) ÷ actual power. In large samples, however, the situation improves and it turns out that the relative error "small," so that the given lower bounds are "good" also in the sense (i).

The next pertinent question now under investigation is whether the proposed test regions (5.6.2), (5.6.5) and (5.6.8) are (a) unbiased and (b) admissible against all relevant alternatives under the respective situations.

Also under investigation is the question as to how these tests compare with the corresponding likelihood ratio or Type II tests. On this it may be observed here, that, except in the degenerate cases where the two methods lead to the identical test, as, for example, the case $k = 2$ under (ii) where both lead to

Hotelling's T, the likelihood ratio tests have a far more difficult small sample (null) distribution problem to contend with than the proposed test. This is with regard to direct usability of the test. The small sample (nonnull) distribution problem (connected with the question of power) would be quite difficult for both types of test, but more so for the likelihood ratio test than for the other. This rules out direct evaluation of power for both types of test, but, while we have fairly good lower bounds to the power of the three different tests proposed, we do not at the moment know of any such lower bounds to the power of the corresponding likelihood ratio tests.

REFERENCES

[The references given below are by no means exhaustive, being merely such as would just suffice for an understanding of this paper.]

[1] R. C. BOSE AND S. N. ROY, "The distribution of the studentized D^2-statistic," *Sankhyā*, Vol. 4 (1938), pp. 10-38.

[2] R. A. FISHER, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc. London*, Ser. A, Vol. 121 (1928), pp. 654-673.

[3] R. A. FISHER, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238-249.

[4] M. A. GIRSHICK, "On the sampling theory of the roots of determinantal equations," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 203-224.

[5] H. HOTELLING, "The generalization of Student's ratio," *Ann. Math. Stat.*, Vol. 2 (1931), pp. 360-378.

[6] H. HOTELLING, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 321-377.

[7] P. L. HSU, "On the distribution of roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250-258.

[8] E. L. LEHMANN AND C. STEIN, "Most powerful tests of composite hypotheses," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 495-516.

[9] D. N. NANDA, "Distribution of a root of a determinantal equation," *Ann. Math Stat.*, Vol. 19 (1948), pp. 47-57.

[10] S. N. ROY, "p-statistics or some generalizations in analysis of variance appropriate to multivariate problems," *Sankhyā*, Vol. 4 (1939), pp. 381-396.

[11] S. N. ROY, "The sampling distribution of p-statistics and certain allied statistics on the nonnull hypothesis," *Sankhyā*, Vol. 6 (1942), pp. 15-34.

[12] S. N. ROY, "The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null-hypothesis," *Sankhyā*, Vol. 7 (1945), pp. 133-158.

[13] S. N. ROY, "Notes on testing of composite hypotheses-II," *Sankhyā*, Vol. 9 (1948), pp. 19-38.

[14] S. N. ROY, "Univariate and multivariate analysis as problems in testing of composite hypotheses-I," *Sankhya*, Vol. 10 (1950), pp. 29-80.

[15] S. N. ROY, "Lecture notes on multivariate analysis," (unpublished).

ON A CLASS OF PROBLEMS RELATED TO THE RANDOM DIVISION OF AN INTERVAL

By D. A. Darling

Columbia University

Summary. Let $X_1, X_2, \cdots, X_n$ be n independent random variables each distributed uniformly over the interval (0, 1), and let $Y_0, Y_1, \cdots, Y_n$ be the respective lengths of the $n + 1$ segments into which the unit interval is divided by the $\{X_i\}$. A fairly wide class of statistical problems is related to finding the distribution of certain functions of the Y_j; these problems are reviewed in Section 1. The principal result of this paper is the development of a contour integral for the characteristic function (ch. fn.) of the random variable $W_n = \sum_{j=0}^{n} h_j(Y_j)$ for quite arbitrary functions $h_j(x)$, this result being essentially an extension of the classical integrals of Dirichlet. The cases of statistical interest correspond to $h_j(x) = h(x)$, independent of j. There is a fairly extensive literature devoted to studying the distributions for various functions $h(x)$. By applying our method these distributions and others are readily obtained, in a closed form in some instances, and generally in an asymptotic form by applying a steepest descent method to the contour integral.

1. Introduction. The statistical problems mentioned above are divided roughly into two classes: problems related to considerations of the Poisson stochastic process occurring in the study of infectious diseases, traffic flow, etc., and problems pertaining to certain nonparametric tests of the hypothesis that a given set of data came from a hypothetical cumulative distribution function (cdf) $F(x)$, which in turn are related to certain "goodness of fit" problems.

In 1946 Greenwood [6], in connection with a problem in epidemiology, posed the general problem of testing whether a given set of points on the unit interval could have arisen from the independent selection of points X_i described above, or whether the set of intervals Y_j they generate are too nearly equal for this hypothesis to be tenable. He suggested the statistic $W_n = \sum_{j=0}^{n} Y_j^2$ and gave a few properties of its distribution. Later Moran [11] proved that W_n had a limiting Gaussian distribution for $n \to \infty$.

If $U_0, U_1, \cdots, U_n$ are $n + 1$ independent random variables each having the density $\beta e^{-\beta x}$, $\beta > 0$, $x > 0$, and if $s_n = U_0 + \cdots + U_n$ it a well known fact that the joint distribution of $\{U_j/s_n\}$, $j = 0, 1, \cdots, n$ is the same as the joint distribution of $\{Y_j\}$, $j = 0, 1, \cdots, n$, the successive lengths of the intervals into which the unit interval is divided by n random points. This correspondence has been used in studying the Poisson stochastic process (cf. [3] chap. 17) in which the interval between successive occurrences of the phenomenon are the U_j. In Greenwood's example these phenomena were the outbreaks of infectious disease.

Received 6/24/52.

In these examples the statistical problems can be reduced to evaluating the distribution of $W_n = \sum h(Y_j)$. In place of Greenwood's suggestion of $h(x) = x^2$, other suggestions were made (cf. the discussion of [6]). Kendall suggested that $h(x) = |x - 1/(n+1)|$ might be analytically more tractable and Irwin suggested $h(x) = (n+1)^{-1}(x - 1/(n+1))^2$. For an analysis of the distribution properties of the extreme Y_j (or U_j) it suffices to consider an $h(x)$ which is 1 for $\alpha < x < \beta$ and zero otherwise. A variety of problems can be reduced to determining the distribution of $W_n = \sum h(Y_j)$ for $h(x)$ of this form. Garwood [5] studied some extremal properties of the Y_j in connection with the occurrence of traffic vehicles on a highway. Fisher [4] had made a similar use in 1925 on the distribution of an extreme amplitude in a problem in harmonic analysis. Kendall made the suggestion of studying the difference (or quotient) of the largest and smallest Y_j as being a more sensitive test function for the equality of the Y_j than Greenwood's sum of squares.

Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables with the common continuous cdf $F(x)$. Let them be relabeled so that $X_1' < X_2' <, \cdots, < X_n'$ and put $X_0' = -\infty$, $X_{n+1}' = +\infty$. Then, as is well known, the joint distribution of $\{F(X_{j+1}') - F(X_j')\}$, $j = 0, 1, \cdots, n$ is the same as the joint distribution of the $\{Y_j\}$, $j = 0, 1, \cdots, n$. Given a set of n data $x_1, x_2, \cdots, x_n$ arranged in increasing order (with $x_0 = -\infty$, $x_{n+1} = +\infty$) a possible test of the hypothesis H that they came from a population whose cdf is $F(x)$ consists in choosing a function $h(x)$ and rejecting H if $\sum h(F(x_{j+1}) - F(x_j))$ is sufficiently large or sufficiently small. Thus the basic problem is, as before, calculating the distribution of $W_n = \sum h(Y_j)$ for various functions h.

Kimball [7] suggested $h(x) = x^\alpha$, $\alpha > 0$, and gave some partial results for the case $\alpha = 2$. The asymptotic character of W_n for $\alpha = 2$ was later analyzed by Moran [11] who proved W_n has a limiting normal distribution for $n \to \infty$. Sherman [13] treated the case $h(x) = \frac{1}{2}|x - 1/(n+1)|$. It will be noted that these tests are somewhat related to the Kolmogoroff-Smirnov tests (cf. [1]) of the "goodness of fit" criteria. A discussion of the relative merits of these tests seems quite academic in view of the complete lack of information concerning their power.

In the present paper we give a unified treatment of these distributions. In Section 2 we develop a simple formula for the ch. fn. of the random variable $W_n = \sum h_j(Y_j)$ (Theorem 2.1) which is essentially an extension of the Dirichlet integral (Theorem 2.2). In Section 3 we study the joint distribution of the Y_j, finding the joint ch. fn. (Theorem 3.1) and the distribution of $Y_0 + Y_1 + \cdots + Y_\nu$. In Section 4 we put $W_n = \sum h(Y_j)$ and develop a few moments of W_n useful in the subsequent work, and in Section 5 we give the asymptotic distribution of W_n for $h(x) = x^\alpha$, the statistic of Greenwood, Moran and Kimball (Theorem 5.1). In Section 6 we analyze the distribution of Sherman and in Section 7 present two more possible test functions which yield readily to our methods.

In Section 8 we study the random variable $N_n(\alpha, \beta)$, the number of those Y_j satisfying $\alpha < Y_j < \beta$, $j = 0, 1, \cdots, n$. As special cases we obtain the

limiting distributions of the number of intervals of "average" size, "small" size and "large" size (Theorems 8.1, 8.2 and 8.3, respectively) and the joint distribution of the largest and smallest Y_j for finite n (Theorem 8.4).

2. The fundamental formula. Let $Y_0, Y_1, \cdots, Y_n$ be the lengths of the $n+1$ intervals into which the unit interval is divided by n random points. The following theorem is the basis for the subsequent analysis in this paper.

THEOREM 2.1. *Let $f_0(x), f_1(x), \cdots, f_n(x)$ be $n+1$ real-valued functions for which the abscissas of convergence of the corresponding Laplace transforms are all less than c. Then*

$$E(f_0(Y_0)f_1(Y_1)\cdots f_n(Y_n)) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z \prod_{j=0}^{n}\int_0^\infty e^{-r_j z} f_j(r_j)\,dr_j\,dz \tag{2.1}$$

the path of integration being the straight line Re $z = c$ (where Re z denotes the real part of z.)

PROOF. We have

$$E\left(\prod_{j=0}^{n} f_j(Y_j)\right) = n!\int_0^1\int_0^{x_n}\int_0^{x_{n-1}}\cdots\int_0^{x_2}\int_0^{x_1} f_0(x_1)f_1(x_2-x_1)\cdots f_{n-1}(x_n-x_{n-1})f_n(1-x_n)\,dx_1\,dx_2\cdots dx_n \tag{2.2}$$

since the joint distribution of the n random points, when arranged in order, has a uniform density differential $n!\,dx_1\,dx_2\cdots dx_n$ over the simplex $0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n \leqq 1$. The trick in "evaluating" this integral consists in considering the following function

$$F(r) = \int_0^r\int_0^{x_n}\int_0^{x_{n-1}}\cdots\int_0^{x_2}\int_0^{x_1} f_0(x_1)f_1(x_2-x_1)\cdots f_n(x_n-x_{n-1})f_n(r-x_n)\,dx_1\,dx_2\cdots dx_n$$

which we want to evaluate at $r = 1$. But it is clear that written this way $F(r)$ is merely the convolution $f_0 * f_1 * \cdots * f_n(r)$ where $g(x)*h(x) = \int_0^x g(x-t)h(t)\,dt$.

Since Laplace transforms multiply under convolution we obtain

$$\int_0^\infty F(r)e^{-zr}\,dr = \prod_{j=0}^{n}\int_0^\infty e^{-zr_j} f_j(r_j)\,dr_j$$

provided $Re\ z > c$. We now simply apply the complex inversion for the Laplace transform to obtain

$$F(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{zx}\prod_{j=0}^{n}\int_0^\infty e^{-r_j z} f_j(r_j)\,dr_j\,dz,$$

and the theorem follows if we put $x = 1$ and supply the factor $n!$.

It is interesting to note that in (2.1) the value of the integral apparently depends on the value of the $f_j(r)$ for $r > 1$ while in (2.2) it does not. As a matter

of fact the functions may be defined quite arbitrarily for $r > 1$ and not affect the value of (2.1).

THEOREM 2.2 *Let D be the domain in E_n defined by $t_i \geqq 0$, $\sum_{i=1}^{n} t_i \leqq 1$. Then for the $f_i(x)$ as in Theorem* 2.1

$$\iint_D \cdots \int f_0(t_1) f_1(t_2) \cdots f_{n-1}(t_n) f_n(1 - t_1 - t_2 - \cdots - t_n)\, dt_1\, dt_2 \cdots dt_n$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^z \prod_{j=0}^{n} \int_0^\infty e^{-r_j z} f_j(r_j)\, dr_j\, dz.$$

To prove the theorem we merely make the change of variables in (2.2) $t_1 = x_1$ $t_2 = x_2 - x_1, \cdots, t_n = x_n - x_{n-1}$ for which the Jacobian is 1.

Theorem 2.2 is, in a sense, a generalization of the integral of Dirichlet—that is, putting $f_0(x) = x^{\alpha_1 - 1}$, $f_1(x) = x^{\alpha_2 - 1}, \cdots, f_{n-1}(x) = x^{\alpha_n - 1}$, $\alpha_i > 0$, and $f_n(x) = f(x)$ we obtain

$$\begin{aligned} &\iint_D \cdots \int t_1^{\alpha_1 - 1} t_2^{\alpha_2 - 1} \cdots t_n^{\alpha_n - 1} f(1 - \textstyle\sum t_j)\, dt_1\, dt_2 \cdots dt_n \\ &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^z \prod_{j=1}^{n} \int_0^\infty e^{-r_j z} r_j^{\alpha_j - 1}\, dr_j \int_0^\infty e^{-rz} f(r)\, dr\, dz \\ &= \prod_{j=1}^{n} \Gamma(\alpha_j) \int_0^\infty f(r) \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{z(1-r)} z^{-\Sigma\alpha_j}\, dz\, dr \\ &= \frac{\Pi\Gamma(\alpha_j)}{\Gamma(\Sigma\alpha_j)} \int_0^1 (1 - r)^{\Sigma\alpha_j - 1} f(r)\, dr \end{aligned} \tag{2.3}$$

since the inner complex integral in (2.3) is zero if $r > 1$ and is $(1 - r)^{\Sigma\alpha_j - 1} / \Gamma(\sum \alpha_j)$ if $0 \leqq r < 1$. This is the classical Dirichlet integral usually developed through the theory of the Beta functions, (cf. Whittaker and Watson [17]).

3. The joint distribution of the $\{Y_j\}$. By means of Theorem 2.1 we can give certain properties of the joint distribution function of the Y_j, $j = 0, 1, \cdots, n$. For the ch. fn. of the Y_j we have the following theorem.

THEOREM 3.1. *If $t_i \neq t_j$ for $i \neq j$, and if $n \geqq 1$, then,*

$$E(e^{i(t_0 Y_0 + t_1 Y_1 + \cdots + t_n Y_n)}) = n! \sum_{j=0}^{n} \frac{e^{it_j}}{\prod_{k \neq j} (it_j - it_k)}$$

and is defined for other values of the t_i by continuity.

To prove the theorem we put $f_j(Y_j) = e^{it_j Y_j}$ in Theorem 2.1, giving

$$E(e^{i\Sigma t_j Y_j}) = \frac{n!}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^z\, dz}{\prod_{j=0}^{n} (z - it_j)}$$

and if all of the t_j are unequal the integral can be replaced by a contour integral surrounding the simple poles. A simple application of the theory of residues then establishes Theorem 3.1.

If some of the t_j are equal we proceed in the same manner. For instance if

$$t_i = \begin{cases} t, & i = 0, 1, \cdots, \nu - 1 \\ 0, & i = \nu, \nu + 1, \cdots, n \end{cases}$$

we obtain the ch. fn. of $x_\nu = \sum_{j=0}^{\nu-1} Y_j$, the νth smallest ordered observation from n observations taken from a rectangular population. Then

$$E(e^{itx_\nu}) = \frac{n!}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^z\, dz}{z^{n-\nu+1}(z - it)^\nu}$$

which again can be evaluated by residues, albeit somewhat awkwardly since the poles are no longer simple. But the density for x_ν is simple to calculate by considering

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{zs}\, dz}{z^{n-\nu+1}(z - it)^\nu}$$

as the inversion of the product of two Laplace transforms

$$\frac{1}{\Gamma(n - \nu + 1)} \int_0^\infty e^{-sz} s^{n-\nu}\, ds = \frac{1}{z^{n-\nu+1}},$$

$$\frac{1}{\Gamma(\nu)} \int_0^\infty e^{-sz} e^{ist} s^{\nu-1}\, ds = \frac{1}{(z - it)^\nu}.$$

Consequently taking the convolution and putting $x = 1$

$$E(e^{itx_\nu}) = \frac{n!}{\Gamma(n - \nu + 1)\Gamma(\nu)} \int_0^1 e^{ist} s^{\nu-1}(1 - s)^{n-\nu}\, ds$$

and thus the density of x_ν is the Beta function $n!s^{\nu-1}(1 - s)^{n-\nu}/\Gamma(n - \nu + 1)\Gamma(\nu)$ as is well known. Other properties also related to the distribution of order statistics from a uniform distribution which have been proved recently by Malmquist [10] may be treated in a like manner.

An evaluation of the mixed moment $E(\prod_j Y_j^{\alpha_j})$ is, of course, easily given in terms of the Dirichlet integral of the preceding section.

4. The distribution of W_n. The statistical problems mentioned in Section 1 may all be reduced to finding the distribution of $W_n = \sum_{j=0}^n h(Y_j)$ for certain functions $h(x)$.

By putting $f_j(x) = e^{i\xi h(x)}$ in (2.1) we obtain

$$E(e^{i\xi W_n}) = \frac{n!}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^W \left(\int_0^\infty e^{-rW + i\xi h(r)}\, dr \right)^{n+1} dW, \tag{4.1}$$

and from this expression we propose to study the distribution of $W_n = \sum h(Y_j)$.

As a preliminary we find the first two moments of W_n which will prove useful in the work to follow. If $\int_0^\infty h^k(r)\,dr$ is finite it is simple to see the (4.1) can be differentiated k times under the integral sign with respect to $i\xi$. Differentiating once and putting $\xi = 0$ we obtain

$$\mu_1 = E(W_n) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^W W^{-n}\,(n+1)\int_0^\infty e^{-rW}\,h(r)\,dr\,dW$$

$$= (n+1)!\int_0^\infty h(r)\,\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{W(1-r)}W^{-n}\,dW\,dr \tag{4.2}$$

$$= n(n+1)\int_0^1 (1-r)^{n-1}h(r)\,dr.$$

Similarly by differentiating twice and setting $\xi = 0$ we obtain the second moment

$$\mu_2 = E(W_n^2) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^W\left\{\frac{n+1}{W^n}\int_0^\infty e^{-rW}h^2(r)\,dr + \frac{n(n+1}{W^{n-1}}\left(\int_0^\infty e^{-rW}h(r)\,dr\right)^2\right\}dW$$

$$= n(n+1)\int_0^1 (1-r)^{n-1}h^2(r)\,dr + n(n+1)!\int_0^\infty\int_0^\infty h(r_1)h(r_2)\,\frac{1}{2\pi i} \tag{4.3}$$

$$\int_{c-i\infty}^{c+i\infty} e^{W(1-r_1-r_2)}\,\frac{dW}{W^{n-1}}\,dr_1\,dr_2$$

$$= n(n+1)\int_0^1 (1-r)^{n-1}h^2(r)\,dr + n^2(n^2-1)\iint_{\substack{0\le r_1+r_2\le 1\\ r_1\ge 0, r_2\ge 0.}} \cdot(1-r_1-r_2)^{n-2}h(r_1)h(r_2)\,dr_1\,dr_2.$$

From (4.2) and (4.3) we can calculate the variance $\sigma^2 = \mu_2 - \mu_1^2$, and proceeding in a similar fashion we can develop all moments if they exist.

5. The distributions of Greenwood, Moran and Kimball. Greenwood [6] suggested $h(x) = x^2$, and Irwin in the discussion of his paper suggested $h(x) = (n+1)^{-1}(x - 1/(n+1)^2$. Moran [11] later found the limiting distribution of Greenwood's statistic was normal. Kimball [7] proposed $h(x) = x^\alpha$ for $\alpha > 0$ and found some partial results for case $\alpha = 2$.

In this section we find the limiting distribution for the case $h(x) = x^\alpha$. We have the following theorem.

THEOREM 5.1. *The random variable* $W_n = \sum Y_j^\alpha$, $\alpha > 0$, $\alpha \neq 1$, *has a limiting normal distribution with the limiting mean and variance*

$$\mu_n \sim \frac{\Gamma(\alpha+1)}{n^{\alpha-1}},$$

$$\sigma_n^2 \sim \frac{1}{n^{2\alpha-1}}\{\Gamma(2\alpha+1) - (\alpha^2+1)\Gamma^2(\alpha+1)\},$$

respectively, that is,

$$\lim_{n\to\infty} Pr\left\{\frac{W_n - \mu_n}{\sigma_n} < x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt$$

for μ_n *and* σ_n^2 *as above.*

Of course if $\alpha = 0$ or $\alpha = 1$ we have $\sigma_n = 0$ and it might be proper to speak of W_n as having a *degenerate* normal distribution.

This theorem will follow by applying a slight variation of the method of steepest descent to the integral (4.1). The proof is given in a fair amount of detail and will serve as a model for the later distributions whose treatment follows essentially the same pattern and for which we give considerably less detailed proofs.

Substituting $h(x) = x^\alpha$ in (4.2) and (4.3) we obtain for the first two moments

$$\mu_1 = \Gamma(\alpha+1)\frac{\Gamma(n+2)}{\Gamma(n+\alpha+1)},$$

$$\mu_2 = \frac{\Gamma(n+2)}{\Gamma(n+2\alpha+1)}(\Gamma(2\alpha+1) - n\Gamma^2(\alpha+1)).$$

The asymptotic character of these moments is easily obtained through the formula

$$\frac{\Gamma(n)}{\Gamma(n+\beta)} = \frac{1}{n^\beta} - \frac{\beta(\beta-1)}{2n^{\beta+1}} + o\left(\frac{1}{n^{\beta+1}}\right) \quad \beta \geqq 0, \qquad n \to \infty$$

giving

$$\mu_1 = \frac{\Gamma(\alpha+1)}{(n+2)^{\alpha-1}} - \Gamma(\alpha+1)\frac{(\alpha-1)(\alpha-2)}{2(n+2)^\alpha} + o\left(\frac{1}{n^\alpha}\right)$$

$$\mu_2 = \frac{\Gamma^2(\alpha+1)}{(n+2)^{2\alpha-2}} + \frac{1}{(n+2)^{2\alpha-1}}(\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)(2\alpha^2 - 3\alpha + 3)) + o\left(\frac{1}{n^{2\alpha-1}}\right)$$

from which we deduce

$$\mu_n \sim \frac{\Gamma(\alpha+1)}{n^{\alpha-1}}$$

(5.1)

$$\sigma_n^2 \sim \frac{1}{n^{2\alpha-1}}(\Gamma(2\alpha+1) - (\alpha^2+1)\Gamma^2(\alpha+1)).$$

Thus for $\alpha > \frac{1}{2}$, $\sigma_n^2 \to 0$ and for $\alpha < \frac{1}{2}$, $\sigma_n^2 \to \infty$ while in the transitional case $\alpha = \frac{1}{2}$ we have $\sigma_n^2 \to 1 - 5\pi/16$.

Using (4.1) we obtain for the ch. fn. of $W_n = \sum Y_j^a$

$$\varphi_n(\xi) = E(\exp(i\xi W_n)) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^W \left(\int_0^\infty e^{-rW+i\xi r^a}\,dr\right)^{n+1} dW$$

for $c > 0$. Letting $\xi = (n+1)^{\alpha-\frac{1}{2}}t$, $W = (n+1)z$ and shifting the contour parallel with itself we find

$$\varphi_n((n+1)^{\alpha-\frac{1}{2}}t) = \frac{(n+1)!}{(n+1)^{n+1}}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{(n+1)z}z^{-n-1}(B_n(z,t))^{n+1}\,dz \tag{5.2}$$

where

$$B_n(z,t) = (n+1)z\int_0^\infty e^{-r(n+1)z+it(n+1)^{\alpha-\frac{1}{2}}r^\alpha}\,dr. \tag{5.3}$$

Now it will turn out that $(B_n(z,t))^{n+1}$ is, aside from a multiplied factor depending on n and t but independent of z, actually a bounded function approaching a limit as $n \to \infty$ for $|t|$ bounded and z arbitrary. This suggests that relative to the dominant term $e^{(n+1)z}z^{-n-1}$ this factor will cause negligible interference when $n \to \infty$, (cf. Szegö [14], p. 220 who treats an example very similar to this.)

If we write $e^{(n+1)z}z^{-n-1} = e^{(n+1)f(z)}$ then $f(z) = z - \log z$ where $\log z$ is real when z is real and positive. Then since $f'(1) = 0$, $f''(1) = 1$, the saddle point is $z = 1$ with the critical direction parallel to the imaginary axis. Hence in (5.2) we merely take $c = 1$ to get the contour of steepest descent.

Thus we put

$$z = 1 + \frac{iy}{\sqrt{n+1}}, \qquad dz = \frac{idy}{\sqrt{n+1}} \tag{5.4}$$

for y in the domain

$$-(n+1)^\delta < y < (n+1)^\delta, \qquad 0 < \delta < \tfrac{1}{2} \tag{5.5}$$

and the entire integral has its essential contribution in this range—the value of the integral extended over the range complementary to (5.5) becoming negligible as $n \to \infty$ after we have modified $B_n(z,t)$ by a factor independent of z. With the substitution (5.4) we find

$$\begin{aligned}\varphi_n((n+1)^{\alpha-\frac{1}{2}}t) &= \frac{(n+1)!e^{n+1}}{(n+1)^{n+3/2}2\pi}\int_{-(n+1)^\delta}^{(n+1)^\delta} e^{-y^2/2}(B_n(z,t))^{n+1}\,dy(1+o(1)) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-(n+1)^\delta}^{(n+1)^\delta} e^{-y^2/2}(B_n(z,t))^{n+1}\,dy(1+o(1))\end{aligned} \tag{5.6}$$

by using Stirling's formula for $(n+1)!$.

Next we turn to $B_n(z,t)$ as given by (5.3). By standard methods we obtain an asymptotic expansion (cf. Watson [16])

$$B_n(z,t) = (n+1)z\int_0^\infty e^{-r(n+1)z}$$

$$\{1 + it(n+1)^{\alpha-1}r^\alpha + \tfrac{1}{2}(it)^2(n+t)^{2\alpha-1}r^{2\alpha} + \cdots\}\,dr$$

$$= 1 + \frac{it\Gamma(\alpha+1)}{(n+1)^1 z^\alpha} + \tfrac{1}{2}(it)^2\frac{\Gamma(2\alpha+1)}{(n+1)z^{2\alpha}} + o(1/n)$$

and for z as in (5.4) and y in the range (5.5)

$$\frac{1}{z^\alpha} = \left(1 + \frac{iy}{(n+1)^{\frac{1}{2}}}\right)^{-\alpha} = 1 - \frac{i\alpha y}{(n+1)^{\frac{1}{2}}} + o(1/n^{\frac{1}{2}})$$

$$\frac{1}{z^{2\alpha}} = \left(1 + \frac{iy}{(n+1)^{\frac{1}{2}}}\right)^{-2\alpha} = 1 + o(1)$$

so that

$$(n+1)\log B_n(z,t)$$

$$= (n+1)\log\left\{1 + \frac{it\Gamma(\alpha+1)}{(n+1)^{\frac{1}{2}}z^\alpha} + \tfrac{1}{2}(it)^2\frac{\Gamma(2\alpha+1)}{(n+1)z^{2\alpha}} + o(1/n)\right\}$$

$$= it\Gamma(\alpha+1)\frac{(n+1)^{\frac{1}{2}}}{z^\alpha} + \tfrac{1}{2}(it)^2\{\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)\}\frac{1}{z^{2\alpha}} + o(1)$$

$$= (n+1)^{\frac{1}{2}}it\Gamma(\alpha+1) - ty\alpha\Gamma(\alpha+1) + \tfrac{1}{2}(it)^2(\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)) + o(1).$$

Using this estimate in (5.6) we obtain

$$\varphi_n((n+1)^{\alpha-1}t)\exp(-(n+1)^{\frac{1}{2}}it\Gamma(\alpha+1)) = \exp\{-\tfrac{1}{2}t^2(\Gamma(2\alpha+1) - \Gamma(\alpha+1))\}$$

$$\cdot\frac{1}{\sqrt{2\pi}}\int_{-(n+1)^{\frac{1}{2}}}^{(n+1)^{\frac{1}{2}}} e^{-y^2/2 - ty\alpha\Gamma(\alpha+1)}\,dy(1 + o(1))$$

and hence

$$\lim_{n\to\infty} E[\exp(it((n+1)^{\alpha-1}W_n - (n+1)^{\frac{1}{2}}\Gamma(\alpha+1))]$$

$$= \lim_{n\to\infty} E\left(\exp\left(it\frac{W_n - (n+1)^{-\alpha+1}\Gamma(\alpha+1)}{(n+1)^{-\alpha+\frac{1}{2}}}\right)\right)$$

$$= \exp\{-\tfrac{1}{2}t^2(\Gamma(2\alpha+1) - \Gamma^2(\alpha+1))\}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2 - ty\alpha\Gamma(\alpha+1)}\,dy$$

$$= \exp\{-\tfrac{1}{2}t^2(\Gamma(2\alpha+1) - (\alpha^2+1)\Gamma^2(\alpha+1))\}$$

which establishes the theorem, and gives an independent derivation for the asymptotic moments.

6. The distribution of Sherman. To avoid some of the difficulties pertaining to the case $h(x) = x^\alpha$ Sherman [13] considered the case $h(x) = \frac{1}{2}\,|\,x - 1/(n+1)\,|$.

Kendall ([6], discussion) had suggested that such a function might be easier to treat because of the simplification of the geometry of the integration. Sherman gave the distribution of $W_n = \frac{1}{2}\sum \mid Y_j - 1/(n+1) \mid$ and proved it had a limiting Gaussian distribution.

In this section we develop the distribution of W_n using (4.1). Here the inner integration can be performed explicitly and the analysis is much simpler. We have in fact

$$\int_0^\infty e^{-rz+\frac{1}{2}i\xi|r-1/(n+1)|}\,dr = \frac{e^{i\xi/2(n+1)}-e^{-z/(n+1)}}{z+\frac{1}{2}i\xi} + \frac{e^{-z/(n+1)}}{z-\frac{1}{2}i\xi} \tag{6.1}$$

so that using (4.1)

$$\begin{aligned}
\varphi_n(\xi) &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z\left\{\frac{e^{i\xi/2(n+1)}-e^{-z/(n+1)}}{z+\frac{1}{2}i\xi} + \frac{e^{-z/(n+1)}}{z-\frac{1}{2}i\xi}\right\}^{n+1} dz \\
&= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty}\left\{\frac{1}{z-\frac{1}{2}i\xi} + \frac{e^{(n+1)^{-1}(z+\frac{1}{2}i\xi)}-1}{z+\frac{1}{2}i\xi}\right\}^{n+1} dz \\
&= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty}\sum_{j=0}^{n+1}\binom{n+1}{j}\left(\frac{e^{(n+1)^{-1}(z+\frac{1}{2}i\xi)}-1}{z+\frac{1}{2}i\xi}\right)^j \frac{dz}{(z-\frac{1}{2}i\xi)^{n+1-j}} \\
&= n!\sum_{j=1}^{n}\binom{n+1}{j}\frac{1}{(n-j)!}\frac{d^{n-j}}{d(i\xi)^{n-j}}\left(\frac{e^{i\xi/(n+1)}-1}{i\xi}\right)^j
\end{aligned}$$

by a simple application of the theory of residues. From this ch. fn. we can easily deduce the density for W_n.

We rewrite the preceding expression

$$\varphi_n(\xi) = n!\sum_{j=1}^{n}\binom{n+1}{j}\frac{1}{(n-j)!(n+1)^j}\frac{d^{n-j}}{d(i\xi)^{n-j}}\left(\frac{e^{i\xi/(n+1)}-1}{i\xi/(n+1)}\right)^j$$

and invert termwise. Let $X_1, X_2, \cdots$ be independent and uniformly distributed over (0, 1). The density of $X_1 + X_2 + \cdots + X_j$ is then (cf. Cramér [2], p. 245)

$$f_j(x) = \frac{1}{(j-1)!}\sum_{0\le k<x}(-1)^k\binom{j}{k}(x-k)^{j-1} \qquad 0 < x < j. \tag{6.2}$$

Then the density for $1/(n+1)(X_1 + X_2 + \cdots + X_j)$ is $(n+1)f_j((n+1)x)$ and the ch. fn. for it is

$$\left(\frac{e^{i\xi/(n+1)}-1}{i\xi/(n+1)}\right)^j.$$

Hence

$$\frac{d^{n-j}}{d(i\xi)^{n-j}}\left(\frac{e^{i\xi/(n+1)}-1}{i\xi/(n+1)}\right)^j = \int_{-\infty}^{\infty} e^{i\xi x}x^{n-j}(n+1)f_j((n+1)x)\,dx.$$

Having inverted the typical term in $\varphi_n(\xi)$ we obtain for the density of W_n

$$n!\sum_{j=1}^{n}\binom{n+1}{j}\frac{x^{n-j}f_j((n+1)x)}{(n-j)!(n+1)^{j-1}}$$

with $f_j(x)$ as in (6.2).

It is also simple to get an asymptotic distribution for W_n following the pattern of Section 5 exactly. If we put $z = (n+1) + (n+1)^{\frac{1}{2}}iy$, $\xi = (n+1)^{\frac{1}{2}}t$ in (6.1) we obtain after some easy estimates

$$\int_0^\infty e^{-rz+(i\xi/2)|r-1/(n+1)|}\,dr$$

$$= \frac{1}{z}\left\{1 + \frac{ite^{-1}}{(n+1)^{\frac{1}{2}}} + \frac{ty}{n+1}(2e^{-1} - \tfrac{1}{2}) - \frac{t^2}{8(n+1)} + o(1/n)\right\},$$

and we choose the same contour as before with $c = 1$. These same substitutions yield

$$n!e^z\,dz/z^{n+1} = i\sqrt{2\pi}e^{-y^2/2}\,dy(1 + o(1))$$

as in the preceding example so that

$$\varphi_n((n+1)^{\frac{1}{2}}t)e^{-it(n+1)^{\frac{1}{2}}e^{-1}} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2+ty(2e^{-1}-\frac{1}{2})-t^2/8}\,dy$$

$$E(e^{it(n+1)^{\frac{1}{2}}(W_n-e^{-1})}) \to e^{-t^2/2(2e^{-1}-4e^{-2})}$$

which exhibits the approach of W_n to the normal distribution.

7. Other possibilities. If we put $h(x) = \log x$ we can evaluate $\varphi_n(\xi)$ explicitly, obtaining,

$$\varphi_n(\xi) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z\left(\int_0^\infty e^{-rz+i\xi\log r}\,dr\right)^{n+1} dz = \frac{\Gamma(n+1)\Gamma^{n+1}(i\xi+1)}{\Gamma((n+1)(i\xi+1))}.$$

Setting $\xi = (n+1)^{-\frac{1}{2}}t$ and using Stirling's formula we get

$$\log\varphi_n((n+1)^{-\frac{1}{2}}t) = -it(n+1)^{\frac{1}{2}}(\log n + \gamma) - \tfrac{1}{2}t^2(\pi^2/6 - 1) + o(1)$$

and it follows that $\sum \log Y_j$ is asymptotically normally distributed with asymptotic mean and variance $-(n+1)(\log n + \gamma)$ and $(n+1)(\pi^2/6 - 1)$ respectively, γ being Euler's constant, $\gamma = .577\cdots$.

In the preceding examples we have always obtained a limiting normal distribution and it seems a reasonable conjecture in analogy with the central limit theorem that we will generally obtain the asymptotic Gaussian distribution when the two moments (4.2) and (4.3) exist. But it appears very difficult to prove a theorem of this generality. We next give an example for which we do not obtain the normal distribution.

Let $h(x) = 1/x$; then since

$$\int_0^\infty e^{-rz-t/r}\,dr = \frac{1}{z}2\sqrt{\xi z}\,K_1(2\sqrt{\xi z}), \qquad \xi > 0, \quad Re\,z > 0$$

where $K_1(x)$ is the Bessel function, (cf. Watson [16]), we have for the Laplace transform of the density of $W_n = \sum Y_j^{-1}$

$$E(e^{-\xi W_n}) = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^z z^{-n-1}\,(2\sqrt{\xi z}\,K_1(2\sqrt{\xi z}))^{n+1}\,dz.$$

Again letting $z = (n+1) + iy(n+1)^{\frac{1}{2}}$ and $\xi = t(n+1)^{-1}$ we have $2\sqrt{\xi z}K_1(2\sqrt{\xi z}) \to 2\sqrt{t}K_1(2\sqrt{t})$ and this expression is the Laplace transform for the density whose cdf is $e^{-1/x}$, $0 < x < \infty$. It will follow then that $W_n/(n+1)$ has the same limiting distribution as the sum of $(n+1)$ independent random variables each having a cdf $e^{-1/x}$, and thus that this limiting distribution is a quasi-stable law of exponent 1 (cf. Lévy [8], p. 208).

8. The number of intervals satisfying certain inequalities. Let $N_n(\alpha, \beta)$ be the number of those Y_j which satisfy $\alpha < Y_j < \beta$ for $j = 0, 1, \cdots, n$. A number of statistical problems relate to the distribution of $N_n(\alpha, \beta)$ as we have outlined in Section 1.

If we put

$$h(r) = \begin{cases} 1 & \alpha < r < \beta \\ 0 & \text{otherwise} \end{cases}$$

then $N_n(\alpha, \beta) = \sum h(Y_j)$ and our preceding discussion is applicable in studying the distribution of this random variable.

Using (4.1) we have

$$\begin{aligned} E(e^{i\xi N_n(\alpha,\beta)}) &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{z}\left\{\int_0^\infty e^{-rz+i\xi h(r)}\,dr\right\}^{n+1} \\ &= \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{z}z^{-n-1}\{1 + (e^{i\xi}-1)\,(e^{-z\alpha}-e^{-z\beta})\}^{n+1}\,dz \end{aligned} \tag{8.1}$$

and this expression will be a basis for the analysis of N_n.

"Most" of the Y_j are presumably of the order of magnitude $(n+1)^{-1}$, and we first find the asymptotic distribution of the number of those Y_j which lie between $a/(n+1)$ and $b/(n+1)$.

THEOREM 8.1. *The random variable $N_n(a/(n+1), b/(n+1))$ is asymptotically normally distributed with an asymptotic mean and variance*

$$\begin{aligned} \mu_n &\sim (n+1)(e^{-a} - e^{-b}) \\ \sigma_n^2 &\sim (n+1)(e^{-a} - e^{-b} - (ae^{-a} - be^{-b})^2). \end{aligned}$$

The proof parallels the analysis of Section 5. Putting $z = (n+1) + (n+1)^{\frac{1}{2}}iy$, $\xi = (n+1)^{-1}t$ in (8.1) we deduce easily

$$\begin{aligned} &1 + (e^{i\xi}-1)\,(e^{-za/(n+1)} - e^{-zb/(n+1)}) \\ &= 1 + \frac{it}{(n+1)^{\frac{1}{2}}}(e^{-a}-e^{-b}) - \frac{t^2(e^{-a}-e^{-b})}{2(n+1)} + \frac{ty}{n+1}(ae^{-a}-be^{-b}) + o(1/n) \end{aligned}$$

and thus

$$\begin{aligned} &E\left(\exp\left\{i\frac{t}{(n+1)^{\frac{1}{2}}}N_n(a/(n+1), b/(n+1)) - it(e^{-a}-e^{-b})\,(n+1)^{\frac{1}{2}}\right\}\right) \\ &\qquad \to e^{-t^2/2(e^{-a}-e^{-b})}\,\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2+ty(ae^{-a}-be^{-b})}\,dy \end{aligned}$$

$$E\left(\exp\left\{it\,\frac{N_n(a/(n+1),\,b/(n+1))-(n+1)\,(e^{-a}-e^{-b})}{(n+1)^{\frac{1}{2}}}\right\}\right)$$
$$\to e^{-t^2/2(e^{-a}-e^{-b}-(ae^{-a}-be^{-b})^2)},$$

which proves the theorem.

We next analyze the distribution of the number of "small" Y_j. It turns out that with probability 1 only finitely many are of the order of magnitude $(n+1)^{-2}$ as $n \to \infty$.

THEOREM 8.2. *$N_n(a/(n+1)^2,\ b/(n+1)^2)$ has an asymptotic Poisson distribution with parameter $(b-a)$. That is*

$$\lim_{n\to\infty} Pr\{N_n(a/(n+1)^2,\,b/(n+1)^2)=k\} = e^{-(b-a)}\,\frac{(b-a)^k}{k!},\quad k=0,\,1,\,\cdots.$$

To prove the theorem we put $\alpha = a/(n+1)^2$, $\beta = b/(n+1)^2$ and $z = (n+1)+(n+1)^{\frac{1}{2}}iy$ in (8.1), giving

$$(1+(e^{i\xi}-1)(e^{-z\alpha}-e^{-z\gamma}))^{n+1} = e^{(b-a)(e^{i\xi}-1)}(1+o(1)),$$

and we have, arguing as before,

$$E(e^{i\xi N_n(a/(n+1)^2,\,b/(n+1)^2)}) \to e^{(b-a)(e^{i\xi}-1)}$$

which establishes the theorem.

The distribution of the number of "large" Y_j proceeds in a similar way.

THEOREM 8.3. *$N_n((\log(n+1)+a)/(n+1),\ (\log(n+1)+b)/(n+1))$ has an asymptotic Poisson distribution with parameter $(e^{-a}-e^{-b})$;*

$$\lim_{n\to\infty} Pr\left\{N_n\left(\frac{\log(n+1)+a}{n+1},\,\frac{\log(n+1)+b}{n+1}\right)=k\right\}$$
$$= e^{-(e^{-a}-e^{-b})}\,\frac{(e^{-a}-e^{-b})^k}{k!},\qquad k=0,\,1,\,\cdots.$$

Thus only finitely many intervals are as large as $\log n/n$ asymptotically with probability 1. To prove the theorem we put

$$\alpha = \frac{\log\dfrac{n+1}{b}}{n+1},\qquad \beta = \frac{\log\dfrac{n+1}{a}}{n+1}\qquad a<b$$

in (8.1) and take $z = (n+1)+(n+1)^{\frac{1}{2}}iy$ giving

$$(1+(e^{i\xi}-1)(e^{-z\alpha}-e^{-z\beta}))^{n+1} \to e^{(b-a)(e^{i\xi}-1)}$$

and the rest of the proof proceeds as before.

From 8.2 and 8.3 we can find the asymptotic distribution of the largest Y_j and the smallest Y_j. Let, in fact, $U_n = \min(Y_0, Y_1, \cdots, Y_n)$ and $V_n =$

$\max(Y_0, Y_1, \cdots, Y_n)$. Then putting $a = 0, k = 0$ in Theorem 8.2 and $b = \infty$, $k = 0$ in Theorem 8.3 we obtain

$$\lim_{n\to\infty} Pr\{U_n > b/(n+1)^2\} = e^{-b}, \qquad 0 < b < \infty$$

$$\lim_{n\to\infty} Pr\left\{V_n < \frac{\log(n+1) + a}{n+1}\right\} = e^{-e^{-a}}, \qquad -\infty < a < \infty.$$

These two expressions were given by Lévy [9] using geometrical arguments. It is possible to show that U_n and V_n are, besides, asymptotically independent. If we put $\alpha = a/(n+1)^2$ and $\beta = \log\frac{n+1}{b} \Big/ n+1$ in (8.1) and duplicate the above reasoning we get

$$\lim_{n\to\infty} Pr\left\{U_n > a/(n+1)^2, V_n < \frac{\log(n+1) - \log b}{n+1}\right\} = e^{-(a+b)}.$$

However by taking a different attack we can get more precise information about the joint distribution of U_n and V_n.

THEOREM 8.4.

$$Pr\{U_n > \alpha, V_n < \beta\} = Pr\{\alpha < Y_j < \beta, j = 0, 1, \cdots, n\}$$
$$= \sum_{(j)}^{*} \binom{n+1}{j} (-1)^j (1 - \alpha(n+1-j) - \beta j)^n \tag{8.2}$$

where $\sum^*$ *means to include only those terms for which* $1 - \alpha(n+1-j) - \beta j$ *is positive,* $j = 0, 1, \cdots$.

The required probability is clearly the probability that $N_n(\alpha, \beta)$ is equal to $(n+1)$. Hence in (8.1) if we expand the factor in braces and select the coefficient of $e^{i\xi(n+1)}$ we get

$$Pr\{N_n(\alpha,\beta) = n+1\} = \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{z} z^{-n-1} (e^{-z\alpha} - e^{-z\beta})^{n+1}\, dz$$
$$= \sum_{j=0}^{n} \binom{n+1}{j} (-1)^j \frac{n!}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{z(1-\alpha(n+1-j)-\beta j)} z^{-n-1}\, dz$$

and this is equal to (8.2) by a direct application of the residue theorem.

Putting $\alpha = 0$ we obtain the probability that all intervals Y_j are less than β

$$Pr\{V_n < \beta\} = \sum_{0 \le j < 1/\beta} \binom{n+1}{j} (-1)^j (1 - \beta j)^n$$

a result going back to Whitworth [18] and used by Fisher [4] in studying the significance of the largest amplitude in harmonic analysis, and by Garwood [5] in traffic studies. Setting $\beta = 1$ in (8.2) we have only the term corresponding to $j = 0$ in the series, and the distribution of the minimum of the Y_j becomes

$$Pr\{U_n > \alpha\} = (1 - (n+1)\alpha)^n \qquad \alpha < 1/(n+1)$$

which is also a result of considerable age.

There are also interesting relationships between the distributions of U_n and V_n with the work of Robbins [12] and Votow [15] on the measure of a random set.

By using (8.2) it would be easy to find the distribution of $V_n - U_n$ or V_n/U_n and, as suggested by Kendall ([6], discussion), these might be better statistics to test for the equality of the Y_j than the statistics W_n discussed in Sections 5, 6, and 7 above.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] W. Feller, *An Introduction to Probability Theory*, John Wiley and Sons, 1950.

[4] R. A. Fisher, "Tests of significance in harmonic analysis," *Proc. Roy. Soc. Edinburgh, Sect. A.*, Vol. 125 (1929), pp. 54–59.

[5] F. Garwood, "An application of the theory of probability to the operation of vehicular-controlled traffic signals," *J. Roy. Stat. Soc. Suppl.*, Vol. 7 (1940), pp. 65–77.

[6] M. Greenwood, "The statistical study of infectious diseases," *J. Roy. Stat. Soc.*, Vol. 109 (1946), pp. 85–110.

[7] B. F. Kimball, "Some basic theorems for developing tests of fit for the case of nonparametric probability distribution functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540–548.

[8] P. Lévy, *Theorie de l'Addition des Variables Aleatoires*, Gauthier-Villars, Paris, 1937.

[9] Paul Lévy, "Sur la division d'un segment par des points choisis au hasard," *C. R. Acad. Sci. Paris*, Vol. 208 (1939), pp. 147–149.

[10] S. Malmquist, "On a property of order statistics from a rectangular distribution," *Skand. Aktuarietids.*, Vol. 33 (1950), pp. 214–222.

[11] P. Moran, "The random division of an interval," *J. Roy. Stat. Soc. Suppl.*, Vol. 9 (1947), pp. 92–98.

[12] H. E. Robbins, "On the expected value of two statistics," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 321–323.

[13] B. Sherman, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.

[14] G. Szegö, "Orthogonal polynomials," *Amer. Math. Soc. Colloq. Publ.*, Vol. 23 (1939).

[15] D. F. Votaw, "The probability distribution of the measure of a random set," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 240–244.

[16] G. N. Watson, *Theory of Bessel Functions*, Cambridge University Press, 1944.

[17] E. T. Whittaker and G. N. Watson, *Modern Analysis*, Cambridge University Press, 1927.

[18] W. A. Whitworth, *Choice and Chance*, Cambridge University Press, 1897.

SEQUENTIAL DECISION PROBLEMS FOR PROCESSES WITH CONTINUOUS TIME PARAMETER. TESTING HYPOTHESES[1]

A. Dvoretzky, J. Kiefer and J. Wolfowitz

Hebrew University, Jerusalem, Cornell University and University of California at Los Angeles

Summary. The purpose of the present paper is to contribute to the sequential theory of testing hypotheses about stochastic processes with a continuous parameter (say, t which one may think of as time). Sequential decision problems about such processes seem not to have been treated before. Subsequently we shall treat problems of point and interval estimation and general sequential decision problems for such processes. The results, in addition to their interest per se and their practical importance, also shed light on the corresponding results for discrete stochastic processes. The subjects of sequential analysis and the theory of decision functions were founded by Wald, and we treat our present subjects in the spirit of his approach. The general results of decision theory, such as the complete class theorem, carry over to sequential problems about stochastic processes with continuous time parameter. As specific examples we treat the Wiener and Poisson processes and obtain, for example, the exact power function. (For discrete processes the corresponding known results, due to Wald, are approximations).

1. Introduction. Let $\{x_1(t), t \geqq 0\}$ and $\{x_2(t), t \geqq 0\}$ be two different stochastic processes. The statistician observes continuously, beginning at $t = 0$, a process $\{x(t), t \geqq 0\}$ which is either $\{x_1(t)\}$ or $\{x_2(t)\}$, and wishes to decide, as soon as possible, whether $\{x(t)\}$ is $\{x_1(t)\}$ or $\{x_2(t)\}$. "As soon as possible" means the following here. Let T be the time when he reaches a decision (in general this may be a chance variable and need not be a constant). Let E_iT denote the expected value of T when $\{x(t)\} = \{x_i(t)\}$, $i = 1, 2$. Let α_1, α_2 be two positive constants, $\alpha_1 + \alpha_2 < 1$. Subject to the requirement that the probability of an incorrect decision when $\{x(t)\} = \{x_i(t)\}$ be at most α_i, the problem is to give a procedure for deciding between $\{x_1(t)\}$ and $\{x_2(t)\}$ such that $E_i(T)$ is a minimum for $i = 1, 2$. This is simply the same formulation for stochastic processes with continuous parameter as was originally given by Wald ([3], [4]) for stochastic processes with a discrete parameter.

In this paper we shall limit ourselves to stochastic processes which fulfill the following conditions. For every $t \geqq 0$, $x(t)$ is a sufficient statistic for the process, that is, the conditional distribution of the chance function $x(\tau), 0 \leqq \tau \leqq t$, given $x(t)$, is, with probability one for every t, the same for the processes $\{x_1(t)\}$

Received 11/4/52.

[1] This work was sponsored by the Office of Naval Research under a contract with Columbia University and under a National Bureau of Standards contract with the University of California at Los Angeles.

and $\{x_2(t)\}$. For every t and x, both $x_1(t)$ and $x_2(t)$ have frequency functions, say $f_1(x, t)$ and $f_2(x, t)$, respectively. Let

$$\text{(1.1)} \qquad \text{ł}(t) = \log \frac{f_2(x(t), t)}{f_1(x(t), t)} \qquad (\text{ł}(0) = 0).$$

Finally we postulate that the $\text{ł}(t)$ process is one of stationary independent increments, that is, a) for every positive integral k, every $h > 0$, and every sequence $t_1 < t_2 < \cdots < t_k \leqq t$, $\text{ł}(t + h) - \text{ł}(t)$ is distributed independently of $\text{ł}(t_1), \cdots, \text{ł}(t_k)$; b) the distribution of $\text{ł}(t + h) - \text{ł}(t)$ depends only upon h and not upon t.

Thus our theory will include the following problems: 1) testing hypotheses about the parameter of a continuous Poisson process with stationary independent increments (to be discussed in detail below in Section 3); 2) testing hypotheses about the mean of a Wiener process (to be discussed in detail below in Section 4); 3) testing hypotheses about the value of $p(0 < p < 1)$ in the following process with stationary independent increments (called the negative binomial): the probability that $x(t) = k$ for every nonnegative integer k is

$$\Gamma(t + k)p^t(1 - p)^k/\Gamma(k + 1)\Gamma(t);$$

4) testing hypotheses about the value of $\theta(\theta > 0)$ in the following process with stationary independent increments (called the Gamma process): the probability density of $x(t)$ at $x(x \geqq 0)$ is given by $x^{t-1}e^{-x/\theta}/\Gamma(t)\theta^t$.

In practice it is, of course, impossible to observe without error a sample function of a continuous process such as the Poisson process or the Wiener process. Yet in many cases these processes do constitute an excellent approximation to physical reality. For example, the incidence of mesons on a Geiger counter is generally assumed to follow a Poisson process. If the recording lag and the dead time of the Geiger counter are very small, a physicist could use the present theory to decide between two possible values of meson density. In this case continuous observation means simply exact registration of incidence times. As another example, our method, or a modification of it, may be applied to problems of life testing.

Moreover, there are several distinct advantages of the continuous parameter procedure over the discrete one. These are as follows.

The expected duration of observing the process before reaching a decision about which hypothesis to adopt can obviously only be shortened by allowing continuous observation.

Moreover, there are many cases, notably the Poisson and Wiener processes, in which an exact determination of the optimal procedure is possible in the continuous case, while in the discrete case so far only approximations have been derived. Thus, even when treating the discrete case, the continuous case, which is easier to treat, may be used to derive approximations when the unit of time is small.

There may also be other advantages in special problems. Thus it is seen in

Section 3 that in the continuous Poisson process the solution does not depend, as in the discrete case, on the values of the two parameters λ_1 and λ_2, but only on their ratio λ_2/λ_1.

2. Application of the Wald sequential procedure. Optimum character of the test. A careful examination of the results of [5] and [6] shows that their conclusions in no way require that the processes be discrete in time, and under the assumptions about the processes made in the preceding section the following results hold.

i) Let a and b, $b < 0 < a$, be given numbers, and let us employ the Wald sequential probability ratio test as follows. As long as $\dotplus(t)$ lies between b and a, continue observing $\{x(t)\}$. As soon as $\dotplus(t) \leqq b$, stop observing $\{x(t)\}$ and decide $\{x(t)\} = \{x_1(t)\}$. As soon as $\dotplus(t) \geqq a$, stop observing $\{x(t)\}$ and decide $\{x(t)\} = \{x_2(t)\}$. Let $\alpha_i(a, b)$ be the probability of error and $E_i(T \mid a, b)$ be the expected value of T when $\{x(t)\} = \{x_i(t)\}$, $i = 1, 2$. For any other procedure with respective probabilities of error α_1^* and α_2^* and respective expected values E_1^*T an E_2^*T, we have that $\alpha_i^* \leqq \alpha_i$, $i = 1, 2$ implies $E_i^*T \geqq E_i(T \mid a, b)$, that is, the optimum character of the Wald sequential probability ratio test (with respect to all randomized as well as nonrandomized procedures).

ii) Let c, W_1 and W_2 be positive numbers, and let g_i be the a priori probability that $\{x(t)\} = \{x_i(t)\}$, $i = 1, 2$ (cf. remarks about a priori probability distributions in [5] and [6]). There exist two numbers $a(c, W_1, W_2, g_1, g_2)$ and $b(c, W_1, W_2, g_1, g_2)$ such that, if the statistician continues to observe $\{x(t)\}$ until either $\dotplus(t) \leqq b$ or $\dotplus(t) \geqq a$, and then decides respectively that $\{x(t)\} = \{x_1(t)\}$ or $\{x(t)\} = \{x_2(t)\}$, he will minimize $g_1(\alpha_1 W_1 + cE_1T) + g_2(\alpha_2 W_2 + cE_2T)$ with respect to all possible procedures for deciding between $\{x_1(t)\}$ and $\{x_2(t)\}$, where E_iT is the expected value of T when $\{x(t)\} = \{x_i(t)\}$, $i = 1, 2$. (It is of course assumed that $a \geqq b$, with the equality sign not excluded. Also $\dotplus(0) = 0$. Thus if $a = b$, or a $\leqq 0$ or $b \geqq 0$, the decision will always be made at time $t = 0$.)

It is to be understood that any procedure which the statistician will employ should be such that the quantities α_1, α_2, E_1T, and E_2T will be well defined. The consideration of questions of measurability is a little more involved for our problem than it is in [5] and [6], but because of the assumptions on the processes made in the preceding section it can be carried out without difficulty. We shall therefore omit consideration of such questions.

From the remarks at the end of Section 1 and well known results of sequential analysis (see Stein [2]), it follows that $E_iT^k < \infty$ for any sequential probability ratio test and any positive k.

Other important results of sequential analysis established for discrete processes apply also to the continuous parameter case. For example, let $\{z(t), t \geqq 0\}$ $(z(0) = 0)$, be a process with stationary independent increments. Assume that $Ez(1)$ exists and denote it by h. Suppose that one has any stopping rule, that is, there is defined a positive chance variable T such that the set of chance functions for which $T = t$ is defined only by conditions on $z(\tau)$, $0 \leqq \tau \leqq t$. Then Wald's

equation ([3], [7])

$$Ez(T+) = hE(T) \tag{2.1}$$

holds. Suppose also that $Ee^{uz(1)}$ exists for all real u, and denote it by $\phi(u)$. Then Wald's fundamental identity ([4], p. 159)

$$Ee^{uz(T+)}(\phi(u))^{-T} = 1 \tag{2.2}$$

holds for many stopping rules, including in particular the rule where $T = t$ if $z(t) \geqq a$ or $z(t) \leqq b$, while $b < z(\tau) < a$ for $\tau < t$. Here a and b are constants, $a > 0$, $b < 0$. The simplest way of proving these results is to derive them as immediate consequences of a theorem of J. L. Doob on martingales with a continuous parameter ([1], Chap. VII, Theorem 11.8). For (2.1) the martingale process is $\{z(t) - ht\}$, and for (2.2) the martingale process is $\{e^{uz(t)}(\phi(u))^{-t}\}$. Another, more laborious way, of proving these results is to consider the process $\{z(t)\}$ only at time intervals which are integral multiples of Δ, proceed as in [4] or [7], and then let Δ approach zero. This is, however, a laborious way of proving a special case of the martingale theorem.

3. The Wiener process. Let $\{x_1(t)\}$ and $\{x_2(t)\}$ be Wiener processes ($t \geqq 0$, $x_1(0) = x_2(0) = 0$) each with a variance which without loss of generality we take to be one per unit of time. Let m_1 and m_2 ($m_1 \neq m_2$) be the mean values per unit time of $\{x_1(t)\}$ and $\{x_2(t)\}$, respectively. Thus we have the following situation: $t \geqq 0$ is a continuous (time) parameter. For any a_1, $a_2 (0 < a_1 < a_2)$, $x_i(a_2) - x_i(a_1)$ is normally distributed with mean $m_i(a_2 - a_1) (i = 1, 2)$ and variance $(a_2 - a_1)$. For any integral k and sequence $a_1^1 < a_2^1 \leqq a_1^2 < a_2^2 \leqq \cdots \leqq a_1^k < a_2^k$, the k chance variables $x_i(a_2^j) - x_i(a_1^j), j = 1, \cdots, k, i = 1, 2$, are independently distributed. The statistician observes continuously, beginning at $t = 0$, a process $\{x(t)\}$ which is either $\{x_1(t)\}$ or $\{x_2(t)\}$, and wishes to decide whether $\{x(t)\} = \{x_1(t)\}$ or $\{x(t)\} = \{x_2(t)\}$.

At time t_0 the quantity $x(t_0)$ is sufficient for deciding between $\{x_1(t)\}$ and $\{x_2(t)\}$, that is, it is unnecessary to know the previous history of the process. The likelihood ratio $L(x(t), t))$ at time t is given by

$$L(x(t), t) = \frac{\frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}((x(t)-m_2 t)^2/t)}}{\frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}((x(t)-m_1 t)^2/t)}}.$$

Hence

$$\mathfrak{L}(t) = x(t)\,(m_2 - m_1) - \frac{t}{2}\,(m_2^2 - m_1^2).$$

The sample functions of the $\{x(t)\}$ process are continuous with probability one. We choose a and b, $b < 0 < a$, such that the statistician will continue to observe $\{x(t)\}$ only until $\mathfrak{L}(t) = a$ or $\mathfrak{L}(t) = b$. In the first case he will decide that $\{x(t)\} =$

$\{x_2(t)\}$, in the second case that $\{x(t)\} = \{x_1(t)\}$. We shall now find $\alpha_i(a, b)$, $E_i(T \mid a, b)$, and the distribution function of T. The same problem for the discrete stochastic process when one observes $\{x(t)\}$ only at $t = 1, 2, \cdots$ has been studied by Wald ([3], [4]) who gave, inter alia, approximations for these quantities. An examination of his argument shows that, in his problem, his results are approximate only because he neglects the excess of $\dagger(T)$ over a or b. In our problem this excess is zero with probability one, and Wald's formulae cease to be mere approximations and become exact. Thus we have, for example, ([4], p. 50, equation (3.42))

$$\alpha_1(a, b) = \frac{1 - e^b}{e^a - e^b} \tag{3.1}$$

$$\alpha_2(a, b) = \frac{e^b(e^a - 1)}{e^a - e^b}. \tag{3.2}$$

For any Wiener process with variance one per unit of time, not necessarily either $\{x_1(t)\}$ or $\{x_2(t)\}$, the probability that $\dagger(t)$ will reach b before reaching a is given *exactly* by [4], page 50, equation (3.43). Call this probability H. Then, for any Wiener process with variance one per unit of time, not necessarily $\{x_1(t)\}$ or $\{x_2(t)\}$, $ET = (Hb + (1 - H)a)/h$ ([4], page 53, equation (3.57)). These results can be derived from (2.1) and (2.2) by Wald's methods. Also the density function of T is given exactly by formula (A:194) on page 195 of [4].

In practice one has to find a and b to correspond to given values α_1 and α_2. Solving (3.1) and (3.2) we obtain

$$a = \log \frac{1 - \alpha_2}{\alpha_1}, \tag{3.3}$$

$$b = \log \frac{\alpha_2}{1 - \alpha_1}. \tag{3.4}$$

All of the above results are exact because the excess of $\dagger(T)$ over the boundaries a and b is zero with probability one. For the same reason one may already infer the optimal character of the Wald sequential probability ratio test for testing hypotheses about the mean of a Wiener process from the approximations and heuristic arguments given by Wald on pages 196–199 of [4].

One may raise the question how to test hypotheses about the variance of a Wiener process. However, a scrutiny of the problem shows that from a knowledge of a sample function in any interval, no matter how small, one can, with probability one, determine the variance to any arbitrary accuracy, so that the problem is trivial. For suppose $\{x(t)\}$ is a Wiener process with mean value m and variance v, both per unit of time. Suppose the process has been observed from $t = 0$ to $t = H_0$, where H_0 is any positive number. Let N be any integer which will later approach infinity, and write $t_i = iH_0/N$, $i = 0, 1, \cdots, N$. For any i from 1 to N we have

$$E(x(t_i) - x(t_{i-1}))^2 = v\frac{H_0}{N} + m^2\frac{H_0^2}{N^2}.$$

Now, for $i = 1, \cdots, N$, the chance variables

$$\left\{(x(t_i) - x(t_{i-1}))^2 - v\frac{H_0}{N} - m^2\frac{H_0^2}{N^2}\right\}$$

are identically and independently distributed, with variance of order N^{-2} and fourth moment of order N^{-4}. Hence the fourth moment of

$$\frac{\sum_{i=1}^{N}(x(t_i) - x(t_{i-1}))^2}{H_0} - v - \frac{m^2H_0}{N}$$

is of order N^{-2}. Consequently, for any $\epsilon > 0$ we have that

$$P\left\{\left|\frac{\sum_{i=1}^{N}(x(t_i) - x(t_{i-1}))^2}{H_0} - v - \frac{m^2H_0}{N}\right| > \epsilon\right\} < \frac{C}{\epsilon^4N^2}$$

where C is a suitable constant. Since the series $\sum N^{-2}$ converges it follows immediately from the Borel-Cantelli lemma that $(\sum_{i=1}^{N}(x(t_i) - x(t_{i-1}))^2)/H_0$ converges to v with probability one as $N \to \infty$.

4. The Poisson process. In this section we treat the problem of deciding which of two values given in advance represents the correct mean occurrence time of a Poisson process with stationary independent increments.

The probability that a Poisson process with mean occurrence time λ will result in exactly k occurrences between times $t = 0$ and $t = T$ is

$$\frac{(\lambda T)^k}{k!}e^{-\lambda T} \qquad (k = 0, 1, 2, \cdots). \tag{4.1}$$

Let $H_i(i = 1, 2)$ denote the hypothesis that $\lambda = \lambda_i$, where λ_1 and λ_2 are any two different positive numbers. It is clear that the two corresponding processes satisfy the conditions imposed in the introduction. Hence, given two positive numbers α_1, α_2, $(\alpha_1 + \alpha_2 < 1)$, the optimal test procedure for deciding between H_1 and H_2 which satisfies the condition that the probability of a wrong decision when H_i is true does not exceed $\alpha_i(i = 1, 2)$ is given by a Wald sequential probability ratio test.

More specifically, in view of (4.1) we have

$$\mathfrak{z}(t) = x(t)\log\frac{\lambda_2}{\lambda_1} + (\lambda_2 - \lambda_1)t. \tag{4.2}$$

Thus, assuming $\lambda_2 > \lambda_1$, the best decision rule is specified by two numbers $a, b(b < 0 < a)$ in the manner described in the introduction.

Suppose now that α_1 and α_2 are the actual probabilities of error. According to Wald ([4], p. 196) we have

$$\frac{1 - \alpha_2}{\alpha_1} = \frac{P_2(H_2)}{P_1(H_2)}, \qquad \frac{\alpha_2}{1 - \alpha_1} = \frac{P_2(H_1)}{P_1(H_1)}, \tag{4.3}$$

where $P_i(H_j)$ is the probability that hypothesis H_j is accepted when hypothesis H_i is true. By the argument used by Wald we have

$$e^{\inf_i \dagger(T)} = \inf_i e^{\dagger(T)} \leqq \frac{P_2(H_i)}{P_1(H_i)} \leqq \sup_i e^{\dagger(T)} = e^{\sup_i \dagger(T)} \tag{4.4}$$

the $\sup_i$ and $\inf_i$ being taken over all values of $\dagger(T)$ where the observation is stopped at time T with the decision to adopt H_i. In our case we know that if the decision to accept H_2 is adopted at time T we must have $\dagger(T) \geqq a$, while $\dagger(t) < a$ for $t < T$. Since (see (4.2)) $\dagger(t + 0) - \dagger(t) \leqq \log \lambda_2/\lambda_1$ with probability 1 we have from (4.3) and (4.4)

$$e^a \leqq \frac{1 - \alpha_2}{\alpha_1} \leqq \frac{\lambda_2}{\lambda_1} e^a. \tag{4.5}$$

Similarly if at time T we decide to terminate observation and adopt H_1 we must have $\dagger(T) \leqq b$ and $\dagger(t) > b$ for $t < T$. Since with probability 1 we have $\dagger(t) \geqq \dagger(t - 0)$ we find that $\dagger(T) = b$ with probability 1. Therefore

$$\frac{\alpha_2}{1 - \alpha_1} = e^b. \tag{4.6}$$

We see here one of the advantages of continuous observation over observation at discrete times only. If we were treating the problem in the conventional manner we would have (4.6) replaced by an inequality, while only the first of the inequalities (4.5) could be derived in the above manner.

Thus we have

$$b = \log \frac{\alpha_2}{1 - \alpha_1} \tag{4.7}$$

and

$$\log \frac{\lambda_1}{\lambda_2} + \log \frac{1 - \alpha_2}{\alpha_1} \leqq a \leqq \log \frac{1 - \alpha_2}{\alpha_1}. \tag{4.8}$$

We now proceed to give a method for the exact computation of a. Without additional effort we shall also find the power function of the test.

We put

$$R(t) = \frac{\dagger(t)}{\log \dfrac{\lambda_2}{\lambda_1}} = x(t) - ct \tag{4.9}$$

where $c = (\lambda_2 - \lambda_1)/\log(\lambda_2/\lambda_1)$. Together with the process $\{x(t)\}$ we have to consider also processes differing from it by a constant; that is, we consider processes with arbitrary $x(0)$.

For given a and b, let $V_\lambda(r)$ be the probability that the procedure described above will terminate with the adoption of H_2 when the Poisson parameter is

really λ and $R(0) = r$. We then have

$$R(\Delta t) = \begin{cases} r - c\Delta t \\ r + 1 - c\Delta t \text{ with probability } \begin{cases} 1 - \lambda\Delta t + o(\Delta t) \\ \lambda\Delta t + o(\Delta t) \\ o(\Delta t) \end{cases} \\ \text{any other value} \end{cases}$$

where the $o(\Delta t)$ terms are all smaller than $\lambda^2\Delta t^2$ for $0 < \Delta t < 1/\lambda$.

Putting

$$K = \frac{b}{\log \frac{\lambda_2}{\lambda_1}}, \qquad J = \frac{a}{\log \frac{\lambda_2}{\lambda_1}} \tag{4.10}$$

we have

$$V_\lambda(r) = 0 \qquad \text{for } r \leqq K$$

$$V_\lambda(r) = 1 \qquad \text{for } r \geqq J$$

while for $K < r < J$ we have

$$V_\lambda(r) = (1 - \lambda\Delta t)V_\lambda(r - c\Delta t) + \lambda\Delta t V_\lambda(r + 1 - c\Delta t) + o(\Delta t) \tag{4.11}$$

with $| o(\Delta t) | < \lambda^2(\Delta t)^2$ for $0 < \Delta t < 1/\lambda$. It follows at once that $V_\lambda(r)$ is continuous for $K \leqq r < J$. (It will be discontinuous at $r = J$.) Rewriting (4.11) as

$$\frac{V_\lambda(r) - V_\lambda(r - c\Delta t)}{\Delta t} = -\lambda V_\lambda(r - c\Delta t) + \lambda V_\lambda(r + 1 - c\Delta t) + \frac{o(\Delta t)}{\Delta t} \tag{4.12}$$

and letting $\Delta t \to 0$ we see that $V_\lambda(r)$ is differentiable in the interval $K < r < J$ with the exception of the point $r = J - 1$ (in case $K < J - 1$). Thus we have the difference-differential equation

$$cV'_\lambda(r) + \lambda V_\lambda(r) = \lambda V_\lambda(r + 1) \tag{4.13}$$

for $K < r < J$ and $r \neq J - 1$. The unique solution in $K < r < J$ is determined by the conditions: (i) $V_\lambda(r)$ continuous for $\lambda < J$, (ii) $V_\lambda(K) = 0$, (iii) $V_\lambda(r) = 1$ for $r \geqq J$.

Let $n(r)$ be the integer such that

$$J - r - 1 \leqq n(r) < J - r. \tag{4.14}$$

It is easy to verify that, for $K \leqq r < J$,

$$V_\lambda(r) = 1 + Ce^{-(\lambda/c)r} \sum_{i=0}^{n(r)} \frac{(-1)^i}{i!} \left[(J - r - i) \frac{\lambda}{c} e^{-\lambda/c} \right]^i \tag{4.15}$$

satisfies (4.13) and (i) for every choice of the constant of integration C. To satisfy also (ii) one has merely to choose

$$C = -e^{(\lambda/c)K} \Big/ \sum_{i=1}^{n(K)} \frac{(-1)^i}{i!} \left[(J - K - i) \frac{\lambda}{c} e^{-\lambda/c} \right]^i. \tag{4.16}$$

Putting $r = 0$ to represent the start of the actual probability ratio test as used in applications, we have from (4.15) and (4.16) that the "OC" function corresponding to the given values of λ_1, λ_2, a and b is given by, say,

$$g\left(\frac{\lambda}{c}\right) = 1 - e^{(\lambda/c)K} \frac{\sum_{i=0}^{n(0)} \frac{(-1)^i}{i!}\left[(J-i)\frac{\lambda}{c}e^{-\lambda/c}\right]^i}{\sum_{i=0}^{n(K)} \frac{(-1)^i}{i!}\left[(J-K-i)\frac{\lambda}{c}e^{-\lambda/c}\right]^i}. \tag{4.17}$$

(K is not displayed since it is given explicitly by (4.10).) Now J should be determined so that

$$g\left(\frac{\lambda_1}{c}\right) = g\left(\frac{\log\frac{\lambda_2}{\lambda_1}}{\frac{\lambda_2}{\lambda_1}-1}\right) = \alpha_1;$$

$$g\left(\frac{\lambda_2}{c}\right) = g\left(\frac{\log\frac{\lambda_2}{\lambda_1}}{1-\frac{\lambda_1}{\lambda_2}}\right) = 1 - \alpha_2. \tag{4.18}$$

Each of the equations (4.18) follows from the other and either one may be used to find J.

It should be noticed that the dependence of K and J on λ_1 and λ_2 is only through the ratio λ_2/λ_1. This follows from (4.10) and (4.17) and could also have been foreseen from the nature of the problem. This remark is useful in the numerical tabulation of the values of J and K or, equivalently, of a and b. (The fact that the λ_i are involved only through their ratio is due to the fact that they are not attached to a given time-unit. In the discrete parameter problem there is an absolute unit of time and hence the two λ_i enter as two parameters. The simplification mentioned above therefore does not occur.)

We now derive, in a manner similar to that used above, an expression for the moment generating function $M_\lambda(u; r) = Ee^{uT}$ of the observation time T necessary to reach a decision when $R(0) = r$ and the true Poisson parameter is λ. From a result of C. Stein [2] it follows that for given J, K and λ there is a positive number $u_0 = u_0(J, K, \lambda)$ such that $M_\lambda(u; r)$ is analytic and uniformly bounded in r for each complex u with real part smaller than u_0. By definition we have $M_\lambda(u; r) = 1$ for $r \leqq K$ or $r \geqq J$. In the same way as (4.11) was derived we obtain (for each u with real part smaller than u_0) for $K < r \leqq J - 1$

$$\begin{aligned} M_\lambda(u; r) &= (1 - \lambda\Delta t)E\{e^{ut} \mid R(0) = r, \quad R(\Delta t) = r - c\Delta t\} \\ &\quad + \lambda\Delta t E\{e^{ut} \mid R(0) = r, \quad R(\Delta t) = r + 1 - c\Delta t\} + o(\Delta t) \\ &= (1 - \lambda\Delta t)e^{u\Delta t}E\{e^{u(t-\Delta t)} \mid R(\Delta t) = r - c\Delta t\} \\ &\quad + \lambda\Delta t E\{e^{ut} \mid R(\Delta t) = r + 1 - c\Delta t\} + o(\Delta t), \end{aligned}$$

or

$$M_\lambda(u;r) = (1-\lambda\Delta t)(1+u\Delta t)M_\lambda(u;r-c\Delta t) + \lambda\Delta t M_\lambda(u;r+1-c\Delta t) + o(\Delta t). \tag{4.19}$$

This form is also valid for $J-1 < r < J$ since $M_\lambda(u;r+1-c\Delta t) = 1 + o(1)$ for $r > J-1$. Since the $o(\Delta t)$ term and $M_\lambda(u;r)$ are uniformly bounded in r we deduce, as in the case of $V_\lambda(r)$, that, considered as a function of r, $M_\lambda(u;r)$ is continuous for $r < J$, possesses a derivative for $K < r < J$ and $r \neq J-1$ and satisfies in the last range the equation

$$c\frac{\partial}{\partial r}M_\lambda(u;r) + (\lambda - u)M_\lambda(u;r) = \lambda M_\lambda(u;r+1). \tag{4.20}$$

It can be verified that the solution of (4.20) satisfying the required boundary conditions is given for $K \leqq r < J$ by

$$\begin{aligned} M_\lambda(u;r) &= \left(\frac{\lambda}{\lambda-u}\right)^{n(r)+1} \\ &+ C(u)e^{-r(\lambda-u)/c}\sum_{i=0}^{n(r)}\frac{(-1)^i}{i!}\left[(J-r-i)\frac{\lambda}{c}e^{-(\lambda-u)/c}\right]^i - \frac{\lambda u}{(\lambda-u)^2} \\ &\cdot e^{(\lambda-u)(J-r-1)/c}\sum_{i=0}^{n(r)-1}\left(\frac{\lambda}{\lambda-u}e^{-(\lambda-u)/c}\right)^i\sum_{j=0}^{i}\frac{(-1)^j}{j!}\left[(J-r-i-1)\frac{\lambda-u}{c}\right]^j \end{aligned} \tag{4.21}$$

with $C(u)$ determined so that $M_\lambda(u;K) \equiv 1$.

Let $Z_\lambda(r)$ be the expected length of time before a final decision is adopted. Then $Z_\lambda(r) = \partial/(\partial u)\, M_\lambda(u;r)\,|_{u=0}$. Since $C(0) = 0$ in (4.21) we obtain, on putting $C'(0) = C'$,

$$\begin{aligned} Z_\lambda(r) &= \frac{n(r)+1}{\lambda} + C'e^{-(\lambda/c)r}\sum_{i=0}^{n(r)}\frac{(-1)^i}{i!}\left[(J-r-i)\frac{\lambda}{c}e^{-\lambda/c}\right]^i \\ &- \frac{1}{\lambda}e^{(\lambda/c)(J-r-1)}\sum_{i=0}^{n(r)-1}e^{-(\lambda/c)i}\sum_{j=0}^{i}\frac{(-1)^j}{j!}\left[(J-r-i-1)\frac{\lambda}{c}\right]^j \end{aligned} \tag{4.22}$$

for $K \leqq r < J$ (of course $Z_\lambda(r) \equiv 0$ outside this range and C' is determined so that $Z_\lambda(K) = 0$).

(One could derive (4.22) without using the moment generating function by establishing the equation

$$cZ'_\lambda(r) + \lambda Z_\lambda(r) = 1 + \lambda Z_\lambda(r+1)$$

for $K < r < J, r \neq J-1$.)

If we write in a more explicit manner $Z_\lambda(r \mid \lambda_1, \lambda_2)$ for $Z_\lambda(r)$ with J and K determined as explained above, it is easily seen that

$$Z_{\alpha\lambda}(r \mid \alpha\lambda_1, \alpha\lambda_2) = \frac{1}{\alpha}Z_\lambda(r \mid \lambda_1, \lambda_2) \tag{4.23}$$

for every positive α.

It is possible to treat the negative binomial process in a manner essentially the same in which we have treated the Poisson process above. A complication is caused by the fact that the probability that the chance variable will exceed one in a small time interval is of the same order of magnitude as the probability that the chance variable will be one.

The authors are obliged to Professor J. L. Doob for several helpful remarks.

REFERENCES

[1] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, 1953.
[2] C. Stein, "A note on cumulative sums," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 498–499.
[3] A. Wald, "Sequential tests of statistical hypotheses, *Ann. Math. Stat.*, Vol. 16 (1945), pp. 117–186.
[4] A. Wald, *Sequential Analysis*, John Wiley and Sons, 1947.
[5] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–329.
[6] A. Wald and J. Wolfowitz, "Bayes solutions of sequential decision problems," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 82–99.
[7] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–230.

EQUIVALENT COMPARISONS OF EXPERIMENTS[1]

By David Blackwell

Stanford University and Howard University

1. Summary. Sherman [8] and Stein [9] have shown that a method given by the author [1] for comparing two experiments is equivalent, for experiments with a finite number of outcomes, to the original method introduced by Bohnenblust, Shapley, and Sherman [4]. A new proof of this result is given, and the restriction to experiments with a finite number of outcomes is removed. A class of weaker comparisons—comparison in k-decision problems—is introduced, in three equivalent forms. For dichotomies, all methods are equivalent, and can be described in terms of errors of the first and second kinds.

2. Introduction. An ordered collection $\alpha = (m_1, \cdots, m_n)$ of probability measures on a Borel field $\mathfrak{B}$ of subsets of a space X will be called an *experiment*. Any pair (α, A), where A is a closed bounded convex subset of n-space corresponds to a decision problem as follows. A point $x \in X$ is selected according to one of the distributions m_i; the statistician observes x and then chooses an action d from a given set D, incurring a loss $L(i, d)$. If we associate with d the vector $w(d) = (L(1, d), \cdots, L(n, d))$, the range of $w(d)$ as d varies over D is the set A associated with the problem. Thus we may replace D by A, and suppose that the statistician chooses a point $a = (a_1, \cdots, a_n) \in A$, incurring loss a_i when m_i is the distribution of x. By using randomized decision procedures we increase A to its convex hull, and for simplicity we suppose A closed and bounded as well as convex.

A decision function in the problem (α, A) is a $\mathfrak{B}$-measurable function f from X into A, specifying for each x the action $a = f(x)$ to be taken when x is observed. When m_i is the distribution of x, the expected loss from f is $v_i(f) = \int a_i(x)\, dm_i(x)$; the vector $v(f) = (v_1(f), \cdots, v_n(f))$ is called the loss vector of f, and the range of $v(f)$ as f varies over all decision functions in the problem (α, A) will be denoted by $B(\alpha, A)$. The set $B(\alpha, A)$ will be a closed, bounded, convex subset of n-space [2].

For two experiments α, β with the same n, following Bohnenblust, Shapley, and Sherman, we say that α is *more informative than* β, written $\alpha \supset \beta$, if for every A we have $B(\alpha, A) \supset B(\beta, A)$, that is if every loss vector attainable in problem (β, A) is also attainable in (α, A). For any experiment $\alpha = (m_1, \cdots, m_n)$, let $p_i(x)$ be the density of m_i with respect to $\sum_1^n m_i$, let $p(x) = [p_1(x), \cdots, p_n(x)]$, and let m_α denote the distribution of $p(x)$ when x has distribution $\sum_1^n m_i/n$. Then m_α is a probability measure defined on the set P of all vectors

Received 11/17/52.

[1] This paper was prepared under an Office of Naval Research Contract.

$p = (p_1, \cdots, p_n)$ with $p_i \geqq 0$ and $\sum_1^n p = 1$, and

$$\int p_i dm_\alpha = 1/n; \tag{1}$$

the center of gravity of m_α is the point $(1/n, \cdots, 1/n)$. The measure m_α is called by Bohnenblust, Shapley, and Sherman the standard measure associated with the experiment α. Their basic results connecting m_α and $\supset$ are summarized as Theorem 1 below (for a proof see [1]).

THEOREM 1. *Every probability measure on P with property* (1) *is the standard measure of some experiment; two experiments α and β have the same standard measure if and only if $B(\alpha, A) = B(\beta, A)$ for all A; $\alpha \supset \beta$ if and only if for every continuous convex function $\phi(p)$ on P, $\int \phi\, dm_\alpha \geqq \int \phi\, dm_\beta$.*

An alternative method of comparing two experiments α, β, introduced by the author [1], can best be described in terms of the concept of *stochastic transformation*. If $\mathfrak{B}$, $\mathfrak{C}$ are Borel fields of subsets of X, Y respectively, a *stochastic transformation* T is a function $Q(x, E)$ defined for all $x \in X$ and $E \in \mathfrak{C}$ which for fixed E is a $\mathfrak{B}$-measurable function of x and for fixed x is a probability measure on $\mathfrak{C}$. For any probability measure m on $\mathfrak{B}$, the function $M(E) = \int Q(x, E)\, dm(x)$ is a probability measure on $\mathfrak{C}$, denoted by Tm. If X, Y are Borel sets in n-space and $\mathfrak{B}$, $\mathfrak{C}$ are the Borel subsets of X, Y, T is called *mean-preserving* if $\int y dQ(x, y) = x$ for all x.

If $\alpha = (m_1, \cdots, m_n)$ and $\beta = (M_1, \cdots, M_n)$ are two experiments, with m_i, M_i defined on Borel fields $\mathfrak{B}$, $\mathfrak{C}$ of X, Y respectively, we shall say that α is *sufficient for* β, written $\alpha > \beta$, if there is a stochastic transformation T with $Tm_i = M_i$ for $i = 1, \cdots, n$. Thus $\alpha > \beta$ means that the statistician, observing the result x of α, can, by selecting y according to $Q(x, E)$, obtain a result equivalent to the result of observing β.

The concept $>$ also has a description in terms of standard measures, summarized in

THEOREM 2. [1]. *$\alpha > \beta$ if and only if there is a mean-preserving stochastic transformation T with $Tm_\beta = m_\alpha$.*

If $\alpha > \beta$ and ϕ is any continuous convex function on P,

$$\begin{aligned} \int \phi\, dm_\alpha &= \int \left(\int \phi(p)\, dQ(q, p) \right) dm_\beta(q) \\ &\geqq \int \phi \left(\int p dQ(q, p) \right) dm_\beta(q) \\ &= \int \phi\, dm_\beta, \end{aligned}$$

so that, from Theorem 1 we obtain

THEOREM 3. *$\alpha \succ \beta$ implies $\alpha \supset \beta$.*

The converse of Theorem 3 has been proved, for experiments with a finite number of outcomes, by Sherman [8] and Stein [9]. In Section 3 we give a new proof of the Sherman-Stein theorem, and in Section 4 extend the theorem to arbitrary experiments.

3. The Sherman-Stein Theorem. If the space X of outcomes of the experiment α is finite, consisting say of $x_1, \cdots, x_N$, then α is characterized by the $n \times N$ Markov matrix $P = \| p_{ij} \|$, where $p_{ij} = m_i(x_j)$, and conversely every Markov matrix can be interpreted as an experiment. For two Markov matrices P, Q with the same n, we write $P \supset Q$, $P \succ Q$ if the corresponding experiments are related by $\supset$, $\succ$ respectively.

THEOREM 4. *If P, Q are $n \times N_1$, $n \times N_2$ Markov matrices with $P \supset Q$, then for every $N_2 \times n$ matrix D there is an $N_1 \times N_2$ Markov matrix M with*

$$\text{Trace } (PMD) \leqq \text{Trace } (QD).$$

PROOF. Let A be the convex hull of the rows of D. The decision function f in problem (Q, A) selecting the jth row of D when j is observed has $v_i(f) = \sum_j q_{ij} d_{ji}$, the ith diagonal element of QD.

Since $P \supset Q$, there is a decision function g in problem (P, A), selecting say $a_j \varepsilon A$ when j is observed, with $v_i(g) = \sum_j p_{ij} a_{ji} = v_i(f)$ for all i. Since $a_j \varepsilon A$, there are nonnegative numbers m_{jk} with $\sum_k m_{jk} = 1$ such that $a_{ji} = \sum_k m_{jk} d_{ki}$ for all i. Thus $v_i(g) = \sum_{j,k} p_{ij} m_{jk} d_{ki}$, which is the ith diagonal element of PMD. It follows that M has not only the property asserted in the theorem but the stronger property that PMD and QD have identical diagonal elements.

THEOREM 5. *$P \succ Q$ if and only if there is a Markov matrix M with $PM = Q$.*

This is simply a restatement of the definition of $\succ$ for the special case $X = (1, \cdots, N_1)$, $Y = (1, \cdots, N_2)$, since a stochastic transformation becomes simply an $N_1 \times N_2$ Markov matrix.

THEOREM 6. (Sherman-Stein theorem). *$P \supset Q$ implies $P \succ Q$.*

PROOF. Consider the function $h(D, M) = \text{Trace } (Q - PM)D$, as M varies over all $N_1 \times N_2$ Markov matrices and D varies over all $N_2 \times n$ matrices with $0 \leqq d_{ki} \leqq 1$ for all k, i. Since h is bilinear and the ranges of D, M are closed, bounded, and convex, h has a saddle point [3], that is there exist D_0, M_0 with $h(D_0, M) \geqq h(D_0, M_0) \geqq h(D, M_0)$ for all D, M. From Theorem 4, there is an M with $h(D_0, M) \leqq 0$, so that $h(D, M_0) \leqq 0$ for all D. Writing $U = Q - PM_0$, we have

$$\text{Trace } (UD) \leqq 0 \text{ for all } D,$$

so that $u_{ik} \leqq 0$ for all i, k. Since U is the difference of two Markov matrices, $\sum_{i,k} u_{ik} = 0$, so that $u_{ik} = 0$ for all i, k and $PM_0 = Q$. Thus by Theorem 5, $P \succ Q$.

An alternative form of the Sherman-Stein theorem is

THEOREM 7. *If m_1 and m_2 are any two probability measures on a finite subset X of n-space such that for every continuous convex ϕ defined on the convex hull of X,*

$\int \phi\, dm_1 \geqq \int \phi\, dm_2$, *then there is a mean-preserving stochastic transformation* T *with* $Tm_2 = m_1$.

From Theorems 1 and 2, Theorem 7 implies Theorem 6. Theorem 7 was proved for $n = 1$ by Hardy, Littlewood, and Polya [6], for $n = 2$ without the restriction that X be finite by the author, and in the form given here by Sherman [8] and Stein [9].

PROOF OF THEOREM 7. From Theorems 1 and 2, Theorem 6 implies Theorem 7 if $X \subset P$ and the common center of gravity of m_1, m_2 is $(1/n, \cdots, 1/n)$, since in this case m_1, m_2 are the standard measures of experiments. Imbedding X in $n + 1$ space and performing an appropriate linear transformation reduces the general case in n-space to that of standard measures in $n + 1$ space and completes the proof.

A direct proof of Theorem 7, using the methods of Theorem 6 and not appealing to Theorems 1 and 2 can be given.

4. Equivalence of $\supset$ and $\succ$. In this section we extend Theorem 7, replacing the requirement that X be finite by the weaker requirement that X be bounded. For any two probability measures m, M on a bounded subset X of n-space, we write $M \supset m$ if for every continuous convex ϕ on the convex hull of X $\int \phi\, dM \geqq \int \phi\, dm$ and $M \succ m$ if there is a mean-preserving stochastic transformation (abbreviated m.p.s.t.) T with $Tm = M$. We shall prove

THEOREM 8. *If* $M \supset m$, *then* $M \succ m$.

The method of proof consists of approximating m, M by measures concentrated on finite sets and using Doob's martingale convergence theorems. We first prove

A. *There exist sequences of measures* m_n, M_n *each concentrated on a finite set, with* $m_N \prec m_{N+1} \subset m \subset M \prec M_{N+1} \prec M_N$ *for all* N, *and for every open set* O

$$m_N(O) \to m(O), \qquad M_N(O) \to M(O) \qquad \text{as } N \to \infty.$$

PROOF. For any n-vector $a = (a_1, \cdots, a_n)$ with integral coordinates, let $C(N, a)$ denote the cube consisting of all $t = (t_1, \cdots, t_n)$ with $2^{-N}a_i \leqq t_i < 2^{-N}(a_i + 1)$, let $Z(N, a)$ be the center of gravity of m on $C(N, a)$ and let m_N assign to $Z(N, a)$ measure $m(C(N, a))$. It is easily verified that m_N has the required properties.

To define M_N, let $Q_N(t, E)$ for $t \in C(N, a)$ concentrate on the 2^n vertices of $C(N, a)$ assigning to vertex $2^{-N}(a_1 + \epsilon_1, \cdots, a_n + \epsilon_n)$, where $\epsilon_i = 0$ or 1, measure $b_1 b_2 \cdots b_n$, where $b_i = 2^{-N}A_i + 1 - t_i$ if $\epsilon_i = 0$ and $b_i = t_i - 2^{-N}a_i$ if $\epsilon_i = 1$. The function $Q_N(t, E)$ is a m.p.s.t. U_N, and if we define $M_N = U_N M$, we have also $M_N = U_N M_{N+1}$, so that M_N has the required properties.

B. *There exist sequences* T_N, V_N, W_N *of* m.p.s.t. *each from a finite set of* n*-space to a finite set of* n*-space with*

$$\text{(a)}\quad m_{N+1} = T_N m_N, \qquad \text{(b)}\quad M_{N-1} = V_N M_N, \qquad \text{(c)}\quad M_N = W_N m_N,$$

and

$$\text{(d)} \quad W_N = V_{N+1}W_{N+1}T_N .$$

PROOF. From A there exist sequences T_N and V_N with properties (a) and (b). Also from A, $m_N \subset M_N$, so that, from Theorem 7, there is a m.p.s.t. Y_N from a finite set to a finite set with $M_N = Y_N m_N$. For $D > N$, write

$$Y_{ND} = V_{N+1} \cdots V_D Y_D T_{D-1} \cdots T_N ,$$

so that

$$Y_{ND} = V_{N+1} Y_{N+1,D} T_N \qquad \text{for } D > N + 1,$$

and

$$M_N = Y_{ND} m_N \qquad \text{for } D > N.$$

Let $D \to \infty$ through a subsequence for which Y_{ND} converges for all N, say to W_N. Then W_N satisfies (c) and (d).

PROOF OF THEOREM 8. We specify the joint distribution of two sequences $x_1, x_2, \cdots, y_1, y_2, \cdots$, of n-dimensional chance variables by

C. For any N, the variables $x_1, \cdots, x_N, y_N, \cdots, y_1$ form a Markov chain in the order written. The distribution of x_1 is m_1 and the conditional distributions of x_{i+1} given x_1, y_N given x_N, and y_{i-1} given y_i, are specified by T_i, W_N, and V_i respectively.

Part (d) of B guarantees that the requirements C are consistent, and Kolmogorov's extension theorem [7] then asserts the existence of $x_1, x_2, \cdots, y_1, y_2, \cdots$, with property C. Parts (a), (b), (c) of B imply that x_N, y_N have distributions m_N, M_N respectively. Also the sequence

$$x_1, x_2, \cdots, \cdots, y_2, y_1$$

forms a martingale [5] in the order written; by Doob's martingale theorem [5], $x_N \to x^*$, $y_N \to y^*$ as $N \to \infty$, and $E(y^* \mid x^*) = x^*$. From A, x^* and y^* have distributions m, M respectively, so that $Q(x, E) = \text{Prob}\{y^* \varepsilon E \mid x^* = x\}$ is a m.p.s.t. T with $Tm = M$. This completes the proof.

5. k-decision problems. In this section we introduce a comparison somewhat weaker than $>$. The following lemma will be useful.

LEMMA. *For any experiment α and any closed, bounded set C with convex hull A, $B(\alpha, A)$ = convex hull of $B(\alpha, C)$.*

PROOF. Since both $B(\alpha, A)$ and $B(\alpha, C)$ are closed and $B(\alpha, A)$ is convex [2], it suffices to show that every $v(f) \varepsilon B(\alpha, A)$ can be approximated by points in the convex hull of $B(\alpha, C)$. We may suppose that f assumes only a finite number of values $a_1, \cdots, a_N$, since every f can be approximated by f's of this kind. Say

$$S_j = \{f(x) = a_j\}, \qquad a_j = \sum_{i=1}^{r} \lambda_{ji} c_i, \lambda_{ji} \geqq 0, \qquad \sum_i \lambda_{ji} = 1.$$

For any $h = (h_1, \cdots, h_N)$, $1 \leqq h_i \leqq r$, define

$$f(h) = c_{h_j} \text{ for } x \in S_j, \lambda(h) = \prod_{j=1}^{r} \lambda_{jh_j}.$$

Then $v(f(h)) \in B(\alpha, C)$, and $\sum_h \lambda(h) v[f(h)]$ has for its sth coordinate

$$\begin{aligned}\sum_h \lambda(h) \sum_j \int_{S_j} c_{h_j s}\, dm_s &= \sum_{h,j} m_s(S_j(c_{h_j s}\, \lambda(h))) \\ &= \sum_j m_s(S_j) \left(\sum_i c_{is} \left(\sum_{h: h_j = i} \lambda(h)\right)\right) \\ &= \sum_j m_s(S_j) \left(\sum_i \lambda_{ij} c_{is}\right) = \sum_j a_{js} m_s(S_j) \\ &= s^{th} \text{ coordinate of } v(f).\end{aligned}$$

This completes the proof.

APPLICATION 1. *Let α be any experiment, let $\mathcal{S} = (S, \cdots, S_k)$ be any partition of X into k disjoint $\mathcal{B}$-measurable sets, let $P(\mathcal{S})$ be the $n \times k$ Markov matrix with $p_{ij} = m_i(S_j)$, let $\mathcal{P}^*_{ak}$ be the range of $P(\mathcal{S})$, and let $\mathcal{P}_{ak}$ be the set of all $n \times k$ Markov matrices P which have the property $\alpha \succ P$. Then $\mathcal{P}_{ak}$ is the convex hull of $\mathcal{P}^*_{ak}$.*

This is the special case of the lemma applied to the experiment α' consisting of nk measures M_{ij} with $M_{ij} = m_i$ for $j = 1, \cdots, k$ and C consisting of the $kn \times k$ Markov matrices $P_1, \cdots, P_k$, where P_j has the jth column identically 1 and the remaining columns identically zero.

APPLICATION 2. *For any experiment α and any closed bounded convex set A which is the convex hull of the set of k points $d_1, \cdots, d_k$, $B(\alpha, A)$ is the range of* diag PD *as P varies over $\mathcal{P}_{ak}$, where* diag U *for any $n \times n$ matrix $U = \| u_{ij} \|$ denotes the n-vector $(u_{11}, u_{22}, \cdots, u_{nn})$ and D is the $k \times n$ matrix whose rows are $d_1, \cdots, d_k$.*

If C consists of $d_1, \cdots, d_k$, and f is any decision function in (α, C), say $S_j = \{f = d_j\}$. Then the sth coordinate of $v(f)$ is

$$\sum_j m_s(S_j) d_{js},$$

so that

$$v(f) = \text{diag } P(\mathcal{S}) D.$$

Thus $B(\alpha, C) =$ range of diag PD as P varies over $\mathcal{P}^*_{ak}$. From the lemma, the convex hull of $B(\alpha, C)$ is $B(\alpha, A)$, and from Application (1) the convex hull of the range of diag PD as P varies over $\mathcal{P}^*_{ak}$ is the range of PD as P varies over $\mathcal{P}_{ak}$.

THEOREM 9. *Let α, β be two experiments with the same n. The following conditions are equivalent:*

(1) $\mathcal{P}_{ak} \supset \mathcal{P}_{\beta k}$

(2) *For every A which is the convex hull of a set of k points, $B(\alpha, A) \supset B(\beta, A)$.*

(3) *For every convex function ϕ on n-space which is the maximum of k linear functions, $\int \phi\, dm_\alpha \geqq \int \phi\, dm_\beta$.*

PROOF. Suppose (1) and let $v \in B(\beta, A)$, where A is the convex hull of $d_1, \cdots, d_k$. Then $v =$ diag PD for some $P \in \mathcal{P}_{\beta k}$.

Since $\mathcal{P}_{\beta k} \subset \mathcal{P}_{\alpha k}$, $v = \text{diag } PD$ for some $P \in \mathcal{P}_{\alpha k}$ and $v \in B(\alpha, A)$. Thus (1) implies (2).

Now suppose (2) and let $P \in \mathcal{P}_{\beta k}$. Then for any closed bounded convex set R, let $v \in B(P, R)$, say $v = v(f)$, where $f(j) = r_j \in R, j = 1, \cdots, k$. Then $v \in B(P, R^*)$, where R^* is the convex hull of $r_1, \cdots, r_k$. Since $B(P, R^*) \subset B(\beta, R^*) \subset B(\alpha, R^*)$, $v \in B(\alpha, R^*)$ and consequently $v \in B(\alpha, R)$. Thus $\alpha \supset P$ for any $P \in \mathcal{P}_{\beta k}$ and, by Theorem 8, $\alpha \succ P$. Since $\mathcal{P}_{\alpha k}$ contains all $n \times k$ Markov matrices P with $\alpha \succ P$, $P \in \mathcal{P}_{\alpha k}$ and $\mathcal{P}_{\beta k} \subset \mathcal{P}_{\alpha k}$. Thus (2) implies (1).

In considering (3), we use the fact that the standard measure m_P of an $n \times k$ Markov matrix P is concentrated on k points, which follows immediately from the definition. Suppose (3), let ϕ be the maximum of any finite set $\mathfrak{L}$ of linear functions, and let $P \in \mathcal{P}_{\beta k}$. There is a ψ, the maximum of k functions in $\mathfrak{L}$, which agrees with ψ on the k points on which m_P is concentrated. Then $\int \phi \, dm_\alpha \geqq \int \psi \, dm_\alpha \geqq \int \psi \, dm_\beta \geqq \int \psi \, dm_P = \int \phi \, dm_P$, so that from Theorems 1 and 8, $\alpha \succ P$. Thus $P \in \mathcal{P}_{\alpha k}$, $\mathcal{P}_{\beta k} \subset \mathcal{P}_{\alpha k}$ and (3) implies (1).

Finally, suppose (1) and let $\phi = \max (L_1, \cdots, L_k)$; say

$$U_j = \{L_j(p) = \phi(p), L_i(p) < \phi(p) \text{ for } i < j\}.$$

If $S_j = \{p(x) \; U_j\}$, $\mathcal{S} = (S_1, \cdots, S_k)$ is a partition of X and the experiment $P = P(\mathcal{S})$ associated with β and $\mathcal{S}$ (see Application 1) has a standard measure m_P with

$$m_P(U_j) = m_\beta(U_j),$$

so that

$$\int \phi \, dm_\beta = \int \phi \, dm_P.$$

Since $P \in \mathcal{P}_{\beta k}$, (1) implies $P \in \mathcal{P}_{\alpha k}$, so that $\int \phi \, dm_\alpha \geqq \int \phi \, dm_P = \int \phi \, dm_\beta$. This completes the proof.

If two experiments α, β with the same n satisfy any of the three equivalent conditions of Theorem 9, we shall say that α is more *informative than β for k-decision problems*, written $\alpha \succ_k \beta$. Condition (2) is the direct analogue of $\supset$, and condition (1) is analogous to $\succ$, since it requires that every experiment with k outcomes producible from β is also producible from α. Clearly $\succ_{k+1}$ implies $\succ_k$, and if $\alpha \succ_k \beta$ for all k, then $\alpha \succ \beta$, since $\alpha \succ_k \beta$ for all k implies $\int \phi \, dm_\alpha \geqq \int \phi \, dm_\beta$ for every ϕ which is the maximum of a finite number of linear functions and hence, by approximation, for every continuous convex ϕ. An alternative statement is: if every experiment with a finite number of outcomes which is producible from β is also producible from α, then β is itself producible from α.

Stein (unpublished paper) has shown that in general $\succ_{k+1}$ is actually stronger than $\succ_k$. For $n = 2$, however, all $\succ_k$ for $k \geqq 2$ are equivalent.

THEOREM 10. *If α and β are two experiments with $n = 2$, then $\alpha \succ_2 \beta$ implies $\alpha \succ \beta$.*

PROOF. For $n = 2$, the standard measures m_α and m_β are defined on the line segment $p_i \geqq 0, p_1 + p_2 = 1$. On this line segment, every function ϕ which is the

maximum of a finite number of linear functions is representable as $\sum a_i\phi_i$, where $a_i > 0$ and each ϕ_i is a maximum of two linear functions. Consequently $\alpha \succ_2 \beta$ implies $\alpha \succ_k \beta$ for all k and hence $\alpha \succ \beta$.

COROLLARY. *Let A be the line segment joining $(0, 1)$ and $(1, 0)$. If $B(\alpha, A) \supset B(\beta, A)$, then $\alpha \succ \beta$.*

PROOF. For any line segment A' in the plane, there is a transformation

$$\text{L:} \quad \begin{aligned} x' &= ax + b \\ y' &= cx + d \end{aligned}$$

with $LA = A'$. Since $LB(\alpha, A) = B(\alpha, LA)$ and similarly for β, we have $B(\alpha, A') \supset B(\beta, A')$, so that $\alpha \succ_2 \beta$ and consequently $\alpha \succ \beta$.

For the A of the corollary, the boundary of the set $B(\alpha, A)$ consists of two curves, joining $(0, 1)$ and $(1, 0)$, one of which is the reflection of the other about $(1/2, 1/2)$. Denote by $f_\alpha(t)$ the minimum of u for which $(t, u) \varepsilon B(\alpha, A)$. Then $\alpha \succ \beta$ if and only if $f_\alpha(t) \leqq f_\beta(t)$ for all t, $0 \leqq t \leqq 1$. The function $f_\alpha(t)$ is a nonincreasing convex function of t, representing the minimum attainable error of the second kind when the error of the first kind is fixed at t. Thus an alternative statement of the corollary is:

α is more informative than β if and only if at every level t the error of the second kind with α is less than or equal to the corresponding error with β.

Since if $\alpha \succ \beta$, then an experiment with n independent observations with α is more informative than the corresponding experiment with β [1] we obtain

THEOREM 11. *If for a sample of size 1 at every level t the probability of an error of the second kind with α does not exceed the corresponding probability for β, then the same is true for every sample size.*

REFERENCES

[1] DAVID BLACKWELL, "Comparison of experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 93–102.

[2] DAVID BLACKWELL, "The range of certain vector integrals," *Proc. Amer. Math. Soc.*, Vol. 2 (1951), pp. 390–395.

[3] H. F. BOHNENBLUST, S. KARLIN, AND L. S. SHAPLEY, "Games with continuous convex payoff," *Contributions to the Theory of Games*, Princeton University Press, 1950, pp. 181–192.

[4] H. F. BOHNENBLUST, L. S. SHAPLEY, AND S. SHERMAN, Unpublished paper.

[5] J. L. DOOB, "Continuous parameter martingales," *Proceedings of the Secondary Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 269–277.

[6] G. H. HARDY, J. E. LITTLEWOOD, AND G. POLYA, *Inequalities*, Cambridge University Press, 1934.

[7] A. N. KOLMOGOROV, "Grundbegriffe der wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik*, No. 3, 1933.

[8] S. SHERMAN, "On a theorem of Hardy, Littlewood, Polya, and Blackwell," *Proc. Nat. Acad. Sci.*, U.S.A., Vol. 37 (1951), pp. 826–831.

[9] CHARLES STEIN, "Notes on the Comparison of Experiments," (mimeographed), University of Chicago, 1951.

TESTING ONE SIMPLE HYPOTHESIS AGAINST ANOTHER

By Lionel Weiss

University of Virginia

1. Summary and introduction. For the problem of testing one simple hypothesis against another, of all tests whose probabilities of incorrectly accepting the first hypothesis and of incorrectly accepting the second hypothesis are bounded from above by given bounds, the familiar Wald sequential probability ratio test gives the smallest expectation of sample size under either hypothesis. In this paper, a "generalized sequential probability ratio test" is introduced which differs from the Wald test only in that the same limits (A, B in the usual notation) are not necessarily used at each stage of the sampling, but at the ith stage A_i and B_i are used, where these numbers are predetermined constants. It is shown that for any given test T, there is a generalized sequential probability ratio test G whose probabilities of incorrectly accepting either hypothesis are no larger than the corresponding probabilities for T, and such that the cumulative distribution function of the number of observations required to come to a decision when using G is never below the corresponding distribution function when using T, under either hypothesis. We may then say that "G is uniformly better than T."

2. Assumptions and notation. In this paper we deal with the problem of testing one simple hypothesis H_1 against another simple hypothesis H_2. We assume that under H_i the chance variable X has a distribution with density function $f_i(x)$. Both $f_1(x)$ and $f_2(x)$ are everywhere bounded and have at most a finite number of discontinuities. We make the test by means of a sequence of independent chance variables $(X_1, X_2, \cdots)$, each having the same distribution (the density function of each is $f_i(x)$ under H_i). We assume that for any n and any finite nonzero c,

$$\iint \cdots \iint \prod_{j=1}^{n} f_i(x_j)\, dx_j \to 0 \qquad \text{as} \quad \Delta c \to 0, \quad \text{for} \quad i = 1, 2,$$

the region of integration being

$$c \leqq \frac{\prod^{n} f_2(x_j)}{\prod^{n} f_1(x_j)} \leqq c + \Delta c.$$

The only tests we shall consider are those not involving randomization, and such that in the space of the first n chance variables the regions where H_1 is accepted and H_2 is accepted are Borel sets, for any n.

We define a "generalized sequential probability ratio test" as follows. There

Received 7/3/52.

are two sequences of predetermined nonnegative constants $(\bar{A}_1, \bar{A}_2, \cdots)$ and $(\bar{B}_1, \bar{B}_2, \cdots)$ such that $\bar{A}_i \geqq \bar{B}_i$ for all i. The value ∞ is not excluded. As long as

$$\bar{B}_n < \prod_{j=1}^{n} f_2(x_j) \Big/ \prod_{j=1}^{n} f_1(x_j) < \bar{A}_n$$

we continue sampling. The first time that this does not happen, we accept H_2 if the upper bound was violated, accept H_1 if the lower bound was violated. If $\bar{A}_m = \bar{B}_m$, while for all $i < m$ we have $\bar{A}_i > \bar{B}_i$, the test is truncated at the mth step. In general, any test is said to be truncated at the mth step if the probability of continuing sampling beyond the mth observation is zero under either hypothesis when using the test.

We use the following notation.

$T{:}D_i(n)$ is the probability that the sample size required to come to a decision is less than or equal to n, when the test T is used and H_i is true.

$T.A_i(n)$ is the region in the space of $(X_1, X_2, \cdots, X_n)$ where we accept H_i when the test T is used. To be in this region, we must have taken an nth observation.

$T{:}A_i$ is the region in the ∞-dimensional space where we accept H_i when using the test T.

$T{:}C(n)$ is the region in the space of $(X_1, X_2, \cdots, X_n)$ where we continue sampling when using the test T.

$P_i(R)$ is the probability of falling in any region R when H_i is true.

$P_i(R \mid S)$ is the conditional probability of falling in R, given that we are in S, when H_i is true.

$Q(X, n)$ is $\prod_{j=1}^{n} f_2(x_j) / \prod_{j=1}^{n} f_1(x_j)$. In specifying that we are using a certain test T, we shall not keep repeating the symbol "T:" throughout an expression, but shall use it once at the beginning of the expression and understand that it modifies everything coming after it, until we reach a symbol denoting another test. Thus if T and T' are two tests, the inequality $T{:}P_1(A_1(m)) + P_1(A_2(m)) > T'{:}P_1(A_1)$ means that the probability of coming to a decision at the mth step when using T and H_1 is true exceeds the probability of accepting H_1 when it is true and T' is used.

3. Existence of a sequence of generalized sequential probability ratio tests uniformly better than a given test in the limit.

THEOREM 1. *If T is any test of H_1 against H_2 such that*

$$\lim_{n \to \infty} T{:}D_i(n) = 1 \qquad \text{for } i = 1, 2,$$

there is a sequence $(G_1, G_2, \cdots)$ of generalized sequential probability ratio tests such that

$$G_j{:}D_i(n) \geqq T{:}D_i(n) \qquad \text{for all } n, \text{ all } j, \text{ and } i = 1, 2; \text{ and}$$

$$\lim_{j \to \infty} G_j{:}P_i(A_i) \geqq T{:}P_i(A_i) \qquad \text{for } i = 1, 2.$$

PROOF. (At certain points in the proof, our statements should really be modified for certain sets of probability zero under both H_1 and H_2. The fact that we have neglected to do this in no way affects the proof.) To prove the theorem, we form a sequence of tests $(T_1, T_2 \cdots)$ as follows. T_j coincides with T until the jth observation. If a jth observation is taken, T_j says accept H_2 if $Q(X, j) \geqq 1$, else accept H_1. Then we have

$$T_j{:}D_i(n) \geqq T{:}D_i(n) \qquad \text{for all } n, \text{ all } j, \text{ and } i = 1, 2; \text{ and}$$

$$\lim_{j\to\infty} T_j{:}P_i(A_i) = T{:}P_i(A_i) \qquad \text{for } i = 1, 2.$$

Now for each j, we will replace T_j by a generalized sequential probability ratio test G_j, such that

$$G_j{:}D_i(n) \geqq T_j{:}D_i(n) \qquad \text{for all } n, \text{ all } j, \text{ and } i = 1, 2; \text{ and} \tag{4.1}$$

$$G_j{:}P_i(A_i) \geqq T_j{:}P_i(A_i) \qquad \text{for all } j \text{ and for } i = 1, 2. \tag{4.2}$$

This, with an obvious application of the Bolzano-Weierstrass theorem, will complete the proof of our theorem. Whenever two tests T and T' stand in the same relation to each other as do G_j and T_j in (4.1) and (4.2) we shall write T^*D^*T'. Thus we can state (4.1) and (4.2) more concisely as $G_j^*D^*T_j$ for all j.

Take any integer j and hold it fixed. Let us assume that for some integer m above 1 but not exceeding j we know that for any given test T_j truncated at the jth step there is a test $T_j'(m)$, also truncated at the jth step and coinciding with T_j before the mth observation, such that $T_j'(m)^*D^*T_j$, and also $T_j'(m)$ has the property $W(m)$ defined as follows:

$$T_j'(m){:}A_1(n) \text{ is given by } Q(X, n) \leqq B_n';$$

and

$$T_j'(m){:}A_2(n) \text{ is given by } Q(X, n) \geqq A_n',$$

for all n between m and j inclusive. We shall then show that all this is true for $m - 1$. Since it is certainly true for $m = j$ (with $A_j' = B_j'$, since a decision must be reached by the jth step), by working back to $m = 1$ we will obtain G_j and thus complete the proof of the theorem.

If in the space of the first $m - 1$ chance variables we consider only those points for which we stop sampling at the $(m - 1)^{st}$ observation, we can always transfer points so that in $T_j{:}A_1(m - 1)$ we have $Q(X, m - 1) \leqq c$, while in $T_j{:}A_2(m - 1)$ we have $Q(X, m - 1) > c$, for some nonnegative c, without making the distribution of the sample size or the probability of accepting a true hypothesis less favorable in any respect. This simply requires the application of the Neyman-Pearson lemma to the set of points under consideration. We shall assume that this is done. Suppose that we then find that there is a number r such that the subset S_I of $T_j{:}C(m - 1)$ where $Q(X, m - 1) > r$ and the subset S_{II} of $T_j{:}A_2(m - 1)$ where $Q(X, m - 1) \leqq r$ are both nonempty. We assume

that $P_1(S_I) > 0$, else we would incorporate S_I into $T_j{:}A_2(m-1)$, which could not make the situation less favorable. Similarly, we may assume $P_2(S_{II}) > 0$, and hence $P_1(S_{II}) > 0$, else we incorporate S_{II} into $T_j{:}A_1(m-1)$. With these assumptions, $c < r < \infty$. Then we can find a number R, with $c < R < \infty$, such that if s_I is the subset of $T_j{:}C(m-1)$ where $Q(X, m-1) > R$, and s_{II} is the subset of $T_j{:}A_2(m-1)$ where $Q(X, m-1) \leqq R$, we have $P_1(s_I) = P_1(s_{II}) > 0$. It is clear that s_I and s_{II} are Borel sets. From now on, when we write X we shall understand the generic point $(x_1, x_2, \cdots, x_{m-1})$ of $m-1$ dimensional space. To each point X_{II} of s_{II} we assign a nonnegative number $r(X_{II})$ as follows: $r(X_{II})$ is the greatest lower bound of the set of numbers v such that $P_1[X \text{ in } s_I \text{ and } Q(X, m-1) \leqq v] \geqq P_1[X \text{ in } s_{II} \text{ and } Q(X, m-1) \leqq Q(X_{II}, m-1)]$.

Now let us assume that we are using a test $T_j'(m)$ defined above, where the acceptance and continuation regions from the first to the $(m-1)^{st}$ observation are given by T_j. We modify this test into $T_j''(m)$ as follows. Transfer s_I into $T_j''(m){:}A_2(m-1)$, s_{II} into $T_j''(m){:}C(m-1)$, and for any X in s_{II}, we act in the future as though $Q(X, m-1)$ were equal to $r(X)$. In all other respects, $T_j''(m)$ coincides with $T_j'(m)$. We shall show that $T_j''(m) {}^*D^* T_j'(m)$. Note that $T_j''(m)$ does not in general have the property $W(m)$.

We need the following lemma. For any given u,

$$P_1[X \text{ in } s_I \text{ and } Q(X, m-1) \leqq u] = P_1[X \text{ in } s_{II} \text{ and } r(X) \leqq u].$$

It clearly suffices to prove the lemma for u between g.l.b. $Q(X, m-1)$ for X in s_I and l.u.b. $Q(X, m-1)$ for X in s_I. The proof of the lemma is given in five short sections.

(1) Given any point X' in s_{II}, we have from the definition of $r(X')$:

$$P_1[X \text{ in } s_I \text{ and } Q(X, m-1) \leqq r(X')] = P_1[X \text{ in } s_{II} \text{ and } Q(X, m-1) \leqq Q(X', m-1)].$$

(2) $P_1[X \text{ in } s_{II} \text{ and } r(X) = u] = 0$. For suppose $r(X') = r(X'') = u$, and $Q(X', m-1) < Q(X'', m-1)$. Then we must have $P_1[X \text{ in } s_{II} \text{ and } Q(X', m-1) \leqq Q(X, m-1) \leqq Q(X'', m-1)] = 0$, else we could not have $r(X') = r(X'')$, by (1). We define

$$Q_1 = \text{g.l.b. } Q(X, m-1) \text{ for } X \text{ in } s_{II} \text{ and } r(X) = u,$$

$$Q_2 = \text{l.u.b. } Q(X, m-1) \text{ for } X \text{ in } s_{II} \text{ and } r(X) = u.$$

Then $P_1[X \text{ in } s_{II} \text{ and } Q_1 \leqq Q(X, m-1) \leqq Q_2] \geqq P_1[X \text{ in } s_{II} \text{ and } r(X) = u]$. Since $P_1[Q(X, m-1) \leqq c]$ is a continuous function of c for c in the open interval $(0, \infty)$, and we can clearly assume that we accept H_2 as soon as $Q(X, n) = \infty$ and accept H_1 as soon as $Q(X, n) = 0$, we have $P_1[X \text{ in } s_{II} \text{ and } Q_1 \leqq Q(X, m-1) \leqq Q_2] = 0$, which proves the first sentence of (2).

(3) $r(X)$ is a nondecreasing function of $Q(X, m-1)$, and therefore if X' is in s_{II} the set of points in s_{II} such that $Q(X, m-1) \leqq Q(X', m-1)$ is the

set of points in s_{II} such that $r(X) \leqq r(X')$ (ignoring sets of probability zero under H_1).

(4) If there is a point X' in s_{II} such that $r(X') = u$, we have $P_1[X \text{ in } s_I \text{ and } Q(X, m-1) \leqq u] = P_1[X \text{ in } s_{II} \text{ and } r(X) \leqq u]$, by (1) and (3). By continuity, the same thing is true if there is a sequence of points $(X_1, X_2, \cdots)$ in s_{II} such that $\lim_{i \to \infty} r(X_i) = u$.

(5) For any u not of the type discussed in (4), we define $B(u) =$ l.u.b. $r(X)$ for all X with $r(X) < u$, $b(u) =$ g.l.b. $r(X)$ for all X with $r(X) > u$. We have

$$P_1[X \text{ in } s_I \text{ and } Q(X, m-1) \leqq B(u)] = P_1[X \text{ in } s_{II} \text{ and } r(X) \leqq B(u)],$$

and

$$P_1[X \text{ in } s_I \text{ and } Q(X, m-1) \leqq b(u)] = P_1[X \text{ in } s_{II} \text{ and } r(X) \leqq b(u)].$$

But

$$P_1[X \text{ in } s_{II} \text{ and } r(X) \leqq B(u)] = P_1[X \text{ in } s_{II} \text{ and } r(X) \leqq b(u)],$$

and therefore our lemma is proved.

First we examine what has occurred when H_1 is true. Since $T_j''(m)$ and $T_j'(m)$ coincide until the $(m-1)^{st}$ observation is taken, we start our investigation at the $(m-1)^{st}$ observation. Also, $T_j''(m){:}A_1(m-1)$ is the same set as $T_j'(m){:}A_1(m-1)$. Since $P_1(s_I) = P_1(s_{II})$, we have $T_j''(m){:}P_1(A_2(m-1)) = T_j'(m){:}P_1(A_2(m-1))$. Now choose any number k between m and j inclusive. We shall show that $T_j''(m){:}P_1(A_i(k)) = T_j'(m){:}P_1(A_i(k))$, $i = 1, 2$. This will complete the proof that $T_j''(m) {*} D {*} T_j'(m)$ when H_1 is true. For any set S, we denote the complement by $\bar{S}$. We have

$$T_j''(m){:}P_1(A_i(k))$$
$$= P_1(\bar{s}_{II} \cdot C(m-1)) P_1(A_i(k) \mid \bar{s}_{II} \cdot C(m-1)) + P_1(s_{II} \cdot C(m-1) \cdot A_i(k)),$$

and

$$T_j'(m){:}P_1(A_i(k))$$
$$= P_1(\bar{s}_I \cdot C(m-1)) P_1(A_i(k) \mid \bar{s}_I \cdot C(m-1)) + P_1(s_I \cdot C(m-1) \cdot A_i(k)).$$

But corresponding terms in the expressions on the right of the two equations are equal to each other, therefore the two left sides are equal. The only terms on the right for which equality is not obvious are $T_j'(m){:}P_1(s_I \cdot C(m-1) \cdot A_i(k))$ and $T_j''(m){:}P_1(s_{II} \cdot C(m-1) \cdot A_i(k))$. These are equal to $T_j'(m){:}P_1(A_i(k) \cdot s_I)$ and $T_j''(m){:}P_1(A_i(k) \cdot s_{II})$ respectively. To show that these two latter expressions are equal to each other, we define $Y_1[u, A_i(k), V]$ to be the probability when H_1 is true of falling in $V{:}A_i(k)$ when we arbitrarily assume that $Q(X, m-1) = u$ and then start sampling, using the test V as though the first observation were the mth, the second were the $(m+1)^{st}$, etc. Clearly, $Y_1[u, A_i(k), T_j'(m)]$ and $Y_1[u, A_i(k), T_j''(m)]$ are equal to each other for all u in the open interval $(0, \infty)$, and are continuous on this interval. We also define $F_1(u, s_I)$ as $P_1[X \text{ in } s_I \text{ and}$

$Q(X, m-1) \leqq u]$, and $G_1(u, s_{II})$ as $P_1[X$ in s_{II} and $r(X) \leqq u]$. We know from the lemma that $F_1(u, s_I) = G_1(u, s_{II})$ identically in u. We can assume that $T_j'(m)$ is such that the sets $[X$ in s_I and $Q(X, m-1) = 0]$ and $[X$ in s_I and $Q(X, m-1) = \infty]$ are empty. Then $F_1(u, s_I)$ is continuous at 0 and ∞. We have

$$T_j'(m):P_1(A_i(k)\cdot s_I) = \int_0^\infty Y_1[u, A_i(k), T_j'(m)]\, dF_1(u, s_I),$$

and

$$T_j''(m):P_1(A_i(k)\cdot s_{II}) = \int_0^\infty Y_1[u, A_i(k), T_j''(m)]\, dG_1(u, s_{II}).$$

From the considerations above, we know that these Stieltjes integrals exist and are equal to each other. Thus we have shown that $T_j''(m):P_1(A_i(k))$ is equal to $T_j'(m):P_1(A_i(k))$ for $i = 1, 2$ and for any k between m and j inclusive.

Now we examine the situation when H_2 is true. Once again, we can start our investigation at the $(m-1)^{st}$ observation. We have $T_j'(m):P_2(C(m-1)) = P_2(s_I) + P_2(\bar{s}_I\cdot C(m-1))$, and $T_j''(m):P_2(C(m-1)) = P_2(s_{II}) + P_2(\bar{s}_{II}\cdot C(m-1))$. But the second terms on the right of these two equalities are equal, while $P_2(s_I) > P_2(s_{II})$, since $P_1(s_I) = P_1(s_{II})$, and in s_I, $Q(X, m-1) > R$, while in s_{II}, $Q(X, m-1) \leqq R$. Now we take any k between m and j inclusive, and examine the expressions

$$T_j'(m):P_2(C(k)) = P_2(s_I\cdot C(k)) + P_2(\bar{s}_I\cdot C(m-1))P_2(C(k) \mid \bar{s}_I\cdot C(m-1)),$$

$$T_j''(m):P_2(C(k)) = P_2(s_{II}\cdot C(k)) + P_2(\bar{s}_{II}\cdot C(m-1))P_2(C(k) \mid \bar{s}_{II}\cdot C(m-1)).$$

The second terms on the right of these two equalities are equal to each other. We investigate the first terms on the right. In a notation that will be recognized by analogy with that already used, we have

$$T_j'(m):P_2(s_I\cdot C(k)) = \int_0^\infty Y_2[u, C(k), T_j'(m)]\, dF_2(u, s_I),$$

$$T_j''(m):P_2(s_{II}\cdot C(k)) = \int_0^\infty Y_2[u, C(k), T_j''(m)]\, dG_2(u, s_{II}).$$

But $dF_2(u, s_I) > dG_2(u, s_{II})$ for all u, because $dF_1(u, s_I) = dG_1(u, s_{II})$, while $dF_2(u, s_I) > R\,dF_1(u, s_I)$ and $dG_2(u, s_{II}) \leqq R\,dG_1(u, s_{II})$ for all u. Also, $Y_2[u, C(k), T_j'(m)] = Y_2[u, C(k), T_j''(m)]$ for all u. Therefore we find that $T_j'(m):P_2(C(k)) > T_j''(m):P_2(C(k))$. To complete the proof that $T_j''(m) * D * T_j'(m)$, we have to show that $T_j''(m):P_2(A_2) \geqq T_j'(m):P_2(A_2)$, or that

$$T_j''(m):P_2(\bar{s}_I\cdot\bar{s}_{II}\cdot A_2) + P_2(s_I) + P_2(s_{II}\cdot A_2)$$

$$\geqq T'_j(m):P_2(\bar{s}_I \cdot \bar{s}_{II} \cdot A_2) + P_2(s_{II}) + P_2(s_I \cdot A_2)$$

or, since the first terms on the two sides of the inequality are equal, that $P_2(s_I) - T'_j(m):P_2(s_I \cdot A_2) \geqq P_2(s_{II}) - T''_j(m):P_2(s_{II} \cdot A_2)$, or

$$\int_0^\infty dF_2(u, s_I) - \int_0^\infty Y_2[u, A_2, T'_j(m)]\, dF_2(u, s_I)$$

$$\geqq \int_0^\infty dG_2(u, s_{II}) - \int_0^\infty Y_2[u, A_2, T''_j(m)]\, dG_2(u, s_{II}),$$

or that

$$\int_0^\infty (1 - Y_2[u, A_2, T'_j(m)])\, dF_2(u, s_I) \geqq \int_0^\infty (1 - Y_2[u, A_2, T''_j(m)])\, dG_2(u, s_{II}),$$

and this last inequality is immediately seen to hold.

By the assumption made above, there is a test $\hat{T}_j(m)$ coinciding with $T''_j(m)$ before the mth observation, having the property $W(m)$, and such that $\hat{T}_j(m)*D*T''_j(m)$. Now we transfer points between $\hat{T}_j(m):A_1(m-1)$ and $\hat{T}_j(m):C(m-1)$ so that after the transfer, for any X left in $\hat{T}_j(m):A_1(m-1)$ we have $Q(X, m-1) \leqq S$, and for any X left in $\hat{T}_j(m):C(m-1)$ we have $Q(X, m-1) > S$, where $0 < S < c$. Then, when S is properly chosen, we can show exactly as above that we can define a test $\hat{T}'_j(m)$, coinciding with $\hat{T}_j(m)$ before the mth observation, such that $\hat{T}'_j(m)*D*\hat{T}_j(m)$. Using the assumption made above again, there is a test $\overline{T}_j(m)$ having the property $W(m)$, coinciding with $\hat{T}'_j(m)$ before the mth observation, and such that $\overline{T}_j(m)*D*\hat{T}'_j(m)$. But then we have $\overline{T}_j(m)*D*T'_j(m)$, and also $\overline{T}_j(m):A_1(m-1)$ is of the form $Q(X, m-1) \leqq S$, $\overline{T}_j(m):A_2(m-1)$ is of the form $Q(X, m-1) \geqq R$, and $\overline{T}_j(m):C(m-1)$ is of the form $S < Q(X, m-1) < R$. Thus the existence of $\overline{T}_j(m)$ shows that if our assumption holds starting from the mth observation, it also holds starting from the $(m-1)^{st}$. Since it holds at the jth observation, the theorem is proved. (Note that we were able to carry out the proof no matter what the acceptance and continuation regions were before the $(m-1)^{st}$ observation).

4. Existence of a generalized sequential probability ratio test uniformly better than a given test.

THEOREM 2. *If T is any test of H_1 against H_2 satisfying the assumptions of Sections 2 and 3, then there is a generalized sequential probability ratio test G such that $G*D*T$.*

PROOF. We start with the sequence of generalized sequential probability ratio tests $(G_1, G_2, \cdots)$ of Theorem 1. From this sequence we can choose a subsequence so that the sequence of $\bar{A}_1$ associated with the subsequence of tests converges (convergence to ∞ is allowed throughout this proof). From this subsequence of tests we choose a second subsequence so that the associated sequence of $\bar{B}_1$ converges. From this second subsequence of tests we choose a third sub-

sequence so that the associated sequence of $\bar{A}_2$ converges. We continue this way in an obvious manner. Then we form a new sequence of tests consisting of the first test in the first subsequence, the second test in the second subsequence, $\cdots$, the ith test in the ith subsequence, $\cdots$. Denote this new sequence by $(S_1, S_2, \cdots)$. Define G to be the generalized sequential probability ratio test given by the two sequences of bounds $(A_1^*, A_2^*, \cdots)$, $(B_1^*, B_2^*, \cdots)$, where $A_i^* = \lim_{j\to\infty} (\bar{A}_i$ associated with $S_j)$, $B_i^* = \lim_{j\to\infty} (\bar{B}_i$ associated with $S_j)$. By our construction, these limits exist. To see that G^*D^*T, it suffices to note that $S_j{:}D_i(n) \geqq T{:}D_i(n)$ for all n, all j, and $i = 1, 2$; and $\lim_{j\to\infty} S_j{:}P_i(A_i) \geqq T{:}P_i(A_i)$ for $i = 1, 2$; and also that for any generalized sequential probability ratio test, the probabilities of falling in the various acceptance and continuation regions under either hypothesis are continuous functions of the associated bounds in the two sequences which characterize the test. (Note that any generalized sequential probability ratio test accepts H_1 as soon as $Q(X, n)$ becomes zero, accepts H_2 as soon as $Q(X, n)$ becomes infinite).

5. Relation of results to decision theory. The relation of the results of this paper to general decision theory is fairly clear. In decision theory we are given a loss function, which we shall assume depends only on the true hypothesis, the hypothesis chosen as correct, and the number of observations required to come to a decision. We shall write this loss function as $W(H, D, N)$, where H is the true situation and can equal either 1 or 2, D is the decision as to which hypothesis is correct and can also equal either 1 or 2, and N is the number of observations required to come to a decision. We also make the following reasonable assumptions about the loss function: $W(1, 2, N) \geqq W(1, 1, N)$ for all N, $W(2, 1, N) \geqq W(2, 2, N)$ for all N, and $W(i, j, N)$ is nondecreasing in N for any fixed i and j. Then the discussion of the previous sections shows that if T is any test, there is a generalized sequential probability ratio test G such that

$$G{:}P_i(W(i, D, N) \leqq w) \geqq T{:}P_i(W(i, D, N) \leqq w) \quad \text{for all } w \text{ and for } i = 1, 2.$$

6. Concluding remarks. The restriction to tests not using randomization that we made above is not necessary. For suppose R is any test, with or without randomization, such that $\lim_{n\to\infty} R{:}D_i(n) = 1$ for $i = 1, 2$. Truncating R at the mth observation in the usual way, we get a test $R(m)$ such that

$$R(j){:}D_i(n) \geqq R{:}D_i(n) \qquad \text{for all } n, \text{ all } j, \text{ and } i = 1, 2; \text{ and}$$

$$\lim_{j\to\infty} R(j){:}P_i(A_i) = R{:}P_i(A_i) \qquad \text{for } i = 1, 2.$$

Theorem 5.1 of [1] tells us that there exists a nonrandomized test $T(j)$ such that $T(j){:}P_k(A_i(n)) = R(j){:}P_k(A_i(n))$ for all n and for $i = 1, 2$, $k = 1, 2$. From this, it is easy to see that Theorems 1 and 2 hold if we consider randomized tests.

Also, the restriction that the density functions be bounded can be dropped, and the results still hold.

Finally, similar results hold in those cases where the observations are not taken one at a time, but in groups of predetermined size.

7. Acknowledgment. The author wishes to thank Dr. Milton Sobel for several suggestions which made this paper more readable.

REFERENCE

[1] A. Dvoretzky, A. Wald, and J. Wolfowitz, "Elimination of randomization in certain statistical decision procedures and zero-sum two-person games," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 1–21.

ON THE EXACT EVALUATION OF THE VARIANCES AND COVARIANCES OF ORDER STATISTICS IN SAMPLES FROM THE EXTREME-VALUE DISTRIBUTION[1]

By Julius Lieblein

National Bureau of Standards

Summary. This paper develops explicit closed formulas for the covariances of order statistics in samples from the extreme-value distribution which involve only tabulated functions. Such results do not appear to have been given previously. They have been used in an investigation of the estimation of extreme-value parameters by means of order statistics which will be presented in a fuller report to be submitted to the National Advisory Committee for Aeronautics.

1. Problem. We are concerned with random samples of size n from the "extreme-value" distribution whose cdf is

$$F(x) = \exp(-e^{-y}), \qquad y = \frac{x - u}{\beta}, \qquad -\infty < x < \infty. \tag{1.1}$$

(This distribution was derived as a limiting form of the distribution of the largest value in a sample by Fisher and Tippett [1] and has been extensively studied by Gumbel (e.g. [4], [5]). However, this paper is not concerned with the extremal properties of this distribution.) If the n values after ordering in size are denoted by

$$x_1, x_2, \cdots, x_n, \qquad x_1 \leqq x_2 \leqq \cdots \leqq x_n,$$

then we seek the second-order moments of the x_i, x_j, namely, the variances σ_i^2 and covariances σ_{ij}. The first moments have been tabulated [6] for samples of $n \leqq 100$.

The second moments involve integrals which at first sight look more difficult than the corresponding ones for the normal distribution, which latter have required a very extensive amount of numerical integration. In this paper a method is shown for evaluating the extreme-value integrals in closed form ((3.10) below) involving only tabulated functions. Thus, the extreme-value distribution is brought into the select circle, which previously included only the normal (at least for $n \leqq 6$—see [3]), exponential, and rectangular distributions, and perhaps some others, for which the second moments of the order statistics can be evaluated explicitly without quadratures.

2. Theory. The density function of the ith order statistic, x_i, from the distribu-

Received 12/4/52.

[1] This paper is based on research sponsored by the National Advisory Committee for Aeronautics.

tion (1.1) is

$$p(x) = \frac{n!}{(i-1)!(n-i)!}[F(x)]^{i-1}[1-F(x)]^{n-i}f(x), \qquad -\infty < x < \infty, \tag{2.1}$$

where $x = x_i$, $f(x) = F'(x)$. The joint d.f. of the ith and jth order statistics x_i, x_j, is

$$p(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1} \cdot [1-F(y)]^{n-j}f(x)f(y), \qquad -\infty < x \leqq y < \infty, \tag{2.2}$$

where $x = x_i$, $y = x_j$, $i < j$, $i, j = 1, 2, \cdots, n$. Without loss of generality, we shall henceforth refer only to the standardized or "reduced" extreme-value distribution, with the parameters $\beta = 1$, $u = 0$,

$$P(y) = \exp(-e^{-y}), \qquad -\infty < y < \infty, \tag{2.3}$$

and denote its variable by y.

From the density functions (2.1) and (2.2) we obtain

$$\begin{aligned} E(y_i^k) &= \frac{n!}{(i-1)!(n-i)!}\int_{-\infty}^{\infty} x^k e^{-(i-1)e^{-x}}(1-e^{-e^{-x}})^{n-i}e^{-x-e^{-x}}\,dx \\ &= \frac{n!}{(i-1)!(n-1)!}\sum_{r=0}^{n-i}(-1)^r C_r^{n-i}\int_{-\infty}^{\infty} x^k e^{-x-(i+r)e^{-x}}\,dx, \end{aligned} \tag{2.4}$$

$$\begin{aligned} E(y_i y_j) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \\ &\quad \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{y} xye^{-y-e^{-y}}e^{-x-ie^{-x}}(e^{-e^{-y}}-e^{-e^{-x}})^{j-i-1}(1-e^{-e^{-y}})^{n-j}\,dx\,dy \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \\ &\quad \cdot \sum_{r=0}^{j-i-1}\sum_{s=0}^{n-j}(-1)^{r+s}C_r^{j-i-1}C_s^{n-j}\phi(i+r, j-i-r+s), \end{aligned} \tag{2.5}$$

where the function ϕ is the double integral

$$\phi(t,u) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} xye^{-x-te^{-x}}e^{-y-ue^{-y}}\,dx\,dy, \qquad t, u > 0 \tag{2.6}$$

whose evaluation is the main point of this paper.

3. Evaluation of the integrals.

3.1 *Variance-type integrals.* These integrals are of the general form

$$g_k(c) = \int_{-\infty}^{\infty} x^k e^{-x-ce^{-x}}\,dx, \qquad c > 0.$$

The evaluation given here is not new, but is presented for completeness.

The change of variable $e^{-x} = v$ gives

$$g_k(c) = \int_0^\infty (-\log v)^k e^{-cv}\, dv,$$

which for k a nonnegative integer

$$\begin{aligned} &= (-1)^k \frac{d^k}{dt^k} \int_0^\infty v^{t-1} e^{-cv}\, dv \Big|_{t=1} \\ &= (-1)^k \frac{d^k}{dt^k} [\Gamma(t) c^{-t}] \Big|_{t=1} \end{aligned} \tag{3.1}$$

The needed first two values are

$$g_1(c) = -\left[\frac{\Gamma'(1)}{c} - \frac{\Gamma(1)}{c} \log c\right] = \frac{1}{c}(\gamma + \log c), \tag{3.2}$$

where $\gamma = -\Gamma'(1)$ is Euler's constant, .5772156649 $\cdots$. Likewise,

$$g_2(c) = \frac{1}{c}\left[\frac{\pi^2}{6} + (\gamma + \log c)^2\right]. \tag{3.3}$$

3.2 *Covariance integrals.* An integration by parts applied to the inner integral in (2.6) with "dv" equal to the exponential factor gives

$$\int_{-\infty}^{v} x e^{-x-te^{-x}}\, dx = t^{-1} y e^{-te^{-y}} - t^{-1} \int_{-\infty}^{v} e^{-te^{-x}}\, dx.$$

Hence from (2.6) and (3.1),

$$t\phi(t, u) = g_2(t + u) - \psi(t, u), \tag{3.4}$$

where

$$\psi(t, u) = \int_{-\infty}^{\infty} y e^{-y-ue^{-y}} \left[\int_{-\infty}^{v} e^{-te^{-x}}\, dx\right] dy. \tag{3.5}$$

The function ψ regarded as a simple integral containing a parameter ($t > 0$) may be differentiated under the integral sign, giving, by (3.2),

$$\frac{\partial \psi}{\partial t} = \frac{1}{t} g_1(t + u) = -\frac{1}{t(t+u)}[\gamma + \log(t + u)], \qquad t, u > 0. \tag{3.6}$$

Before integrating this equation, it is convenient to make the change of variable $w = 1 + (t/u)$. After the substitution integrate (3.6) with respect to w from $w = 2$ to $w = w$, and replace the upper limit w in the resulting expression by its value in terms of t, noting that the corresponding limits for t are $t = u$ to $t = t$. The result is

$$\begin{aligned} u[\psi(t, u) - \psi(u, u)] &= \gamma \log(1 + u/t) + \tfrac{1}{2}[\log(t + u)]^2 \\ &\quad - \log u \log t/u - \gamma \log 2 - \tfrac{1}{2}(\log 2u)^2 - \int_2^{1+t/u} \frac{\log w}{w - 1}\, dw. \end{aligned} \tag{3.7}$$

The integral on the right is immediately expressible in terms of *Spence's integral* (or function)

$$L(1+x) = \int_1^{1+x} \frac{\log w}{w-1}\,dw = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2}\,x^n. \tag{3.8}$$

Several tables of this function are cited in [2]. The most extensive of these is given by F. W. Newman [7] to twelve decimal places.

It remains only to evaluate $\psi(u, u)$. From (2.6) and (3.2),

$$\begin{aligned}\phi(u,u) &= \int_{-\infty}^{\infty} ye^{-y-ue^{-y}} \left(\int_{-\infty}^{y} xe^{-x-ue^{-x}}\,dx\right) dy \\ &= \int_{-\infty}^{\infty} \tfrac{1}{2}\frac{d}{dy}\left(\int_{-\infty}^{y} xe^{-x-ue^{-x}}\,dx\right)^2 dy \\ &= \frac{1}{2u^2}(\gamma + \log u)^2.\end{aligned}$$

This value when substituted in (3.4) gives $\psi(u, u)$. Combining this result with (3.7) and the easily obtainable value $L(2) = \pi^2/12$ gives, after a little algebra, the following formula:

$$\begin{aligned}2tu\,\phi(t,u) = (u-t)g_2(t+u) + t^2[g_1(t)]^2 + 2L\left(1+\frac{t}{u}\right) \\ - \left(\log\frac{t}{u}\right)^2 - \frac{\pi^2}{6},\end{aligned} \tag{3.9}$$

where the functions $g_1(t)$, $g_2(t)$ are given by (3.2), (3.3). This may be simplified a little by use of the following property of Spence's function:

$$L(1+x) + L\left(1+\frac{1}{x}\right) = \tfrac{1}{2}(\log x)^2 + \frac{\pi^2}{6},$$

giving the result

$$2tu\,\phi(t,u) = (u-t)g_2(t+u) + t^2[g_1(t)]^2 - 2L\left(1+\frac{u}{t}\right) + \frac{\pi^2}{6}. \tag{3.10}$$

The above results (3.9), (3.10), together with (2.4), (2.5), make possible the evaluation of all the variances and covariances. This requires the calculation of n values of g_1 and of g_2, and $\frac{1}{2}n(n-1)$ values of ϕ.

The calculation may be simplified with the aid of the relation

$$\phi(t,u) + \phi(u,t) = g_1(t)g_1(u), \tag{3.11}$$

which may be derived from (2.6) and (3.2) by means of a change in the order of integration. Thus (3.10) need be used only for $t \leqq u$, so that (3.11) reduces the number of values of ϕ by almost half, unless n is small, say $n < 10$.

4. Illustration. The above formulas have been used by the author in an investigation of estimation of extreme-value parameters by means of order statistics.

The results of this research, including a table of the first two moments for small samples, will be reported elsewhere.

The following computations for $n = 3$ illustrate the procedure described in this article.

From (3.2),

$$g_1(1) = \gamma = 0.57721\ 57$$
$$g_1(2) = 0.63518\ 14$$
$$g_1(3) = 0.55860\ 93.$$

The means are then given by (2.4) and (3.2) as

$$E(y_1) = 3[g_1(1) - 2g_1(2) + g_1(3)] = -0.40361\ 4$$
$$E(y_2) = 6[g_1(2) - g_1(3)] = +0.45943\ 3$$
$$E(y_3) = 3g_1(3) = +1.67582\ 8.$$

As a simple check, these three values sum to 3γ to within six decimal places, and also agree with those in [6]. (The notation in the table cited differs from that used here: $E(y_i)$ in this paper corresponds to $E(y_{n-i})$ in the table.) Next, from (3.3), we have

$$g_2(1) = \frac{\pi^2}{6} + \gamma^2 = 1.97811\ 2$$
$$g_2(2) = 1.62937\ 8$$
$$g_2(3) = 1.48444\ 4.$$

The mixed function $\phi(t, u)$ is then given by (3.9) and (3.10):

$$\phi(1, 1) = \tfrac{1}{2}(\gamma^2) = 0.16658\ \ 9$$
$$\phi(1, 2) = \tfrac{1}{4}[g_2(3) + \gamma^2 + 2L(1\tfrac{1}{2}) - (1n2)^2 - \pi^2/6] = 0.14726\ \ 6$$
$$\phi(2, 1) = \tfrac{1}{4}\{-g_2(3) + 4[g_1(2)]^2 - 2L(1\tfrac{1}{2}) + \pi^2/6\} = 0.21937\ \ 1.$$

Newman's table [7] provides the value of the function $L(1\frac{1}{2}) = 0.44841\ 42069$. Finally, equations (2.4) and (2.5) give, for the moments about the origin,

$$\| E(y_i y_j) \| = \begin{bmatrix} 0.61140 & 0.11594 & -0.43263 \\ 0.11594 & 0.86960 & 1.31622 \\ -0.43263 & 1.31622 & 4.45333 \end{bmatrix}$$

whence the moments about the mean are given by

$$\| \sigma(y_i y_j) \| = \begin{bmatrix} 0.44850 & 0.30137 & 0.24376 \\ 0.30137 & 0.65852 & 0.54629 \\ 0.24376 & 0.54629 & 1.64493 \end{bmatrix}.$$

The final results are correct to about four decimal places. (One additional place is shown for checking purposes.)

As a check, we should have

$$\sum_{j=1}^{3}\sum_{i=1}^{3}\sigma(y_i y_j) = \sigma^2\left(\sum_{i=1}^{3} y_i\right) = 9\sigma^2(\bar{y}) = 3\sigma_y^2 = \frac{\pi^2}{2},$$

since σ_y^2, the variance of the distribution $P(y)$ in (2.3), is known to be $\pi^2/6$. The left side of this equation is found to be 4.93479; the right side, 4.93480. This type of check cannot be considered to be very effective, however, as only gross errors, and not compensating ones, will ordinarily be revealed.

The reader should be cautioned that, unless n is fairly small, it may be necessary to carry out the calculations to a considerably greater number of places than is desired in the results. This results from the presence of binomial coefficients and alternating signs in formulas (2.4) and (2.5), both of which operate to reduce accuracy rapidly as n increases.

REFERENCES

[1] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Proc. Cambridge Philos. Soc.*, Vol. 24 (1928), pp. 180–190.

[2] A. Fletcher, J. C. P. Miller, and L. Rosenhead, *An Index of Mathematical Tables*, McGraw-Hill Book Company, Inc., 1946, pp. 343–344.

[3] H. J. Godwin, "Some low moments of order statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 279–285.

[4] E. J. Gumbel, "Les valeurs extrêmes des distributions statistiques," *Ann. Inst. H. Poincaré*, Vol. 4 (1935), pp. 115–158.

[5] E. J. Gumbel, "The return period of flood flows," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 163–190.

[6] *Table of the First Moment of Ranked Extremes*, National Bureau of Standards Report 1167, September 20, 1951, (special report submitted to the National Advisory Committee for Aeronautics.)

[7] F. W. Newman, *The Higher Trigonometry, Superrationals of Second Order*, Macmillan and Bowes, Cambridge University Press, (1892), pp. 64–65.

GENERALIZED HIT PROBABILITIES WITH A GAUSSIAN TARGET, II

By D. A. S. Fraser

University of Toronto

1. Summary. In a recent paper [2] the author developed a discrete distribution and several derived limiting distributions for the number of "hits" on a k-dimensional Gaussian target. The purpose of the present paper is to apply these results to the two-dimensional problem considered by Cunningham and Hynd [1]. A general expression and two limiting forms are obtained for the probability of at least one hit. The numerical evaluation using the data in [1] is considered for $n = 5$ rounds, and the probability of at least one hit is plotted in Fig. 1 for various combinations of aiming and dispersion error. For a given over-all time interval the evaluation for large n is discussed in Section 5 and illustrated using the data from [1].

2. Introduction. In 1946 Cunningham and Hynd considered a problem in aerial gunnery: to find the probability of hitting a moving target at least once. The various factors entering into the problem may be described as follows. The point at which the gun is aiming is found to wander back and forth across the target; its successive positions when n rounds are fired can be represented by a multivariate normal distribution with independence between the horizontal and vertical coordinates. We let the coordinates of this point of aim for the ith round be x_{1i}, x_{2i} (the point of aim is called a prediction in [2]). The dispersion error of the gun is also assumed to be normal; we let the trajectory coordinates be y_{1i}, y_{2i}.

In [1] the target was taken to be circular. Here we assume that it is Gaussian diffuse; that is, the probability of a hit is given up to a constant factor by a Gaussian p.d.f. of the coordinates of the trajectory. Because of the irregular outline of a plane and the sharp "drop off" of the p.d.f. proceeding out from the center, this is not an unreasonable assumption.

3. The k-dimensional problem. The k-dimensional problem treated in [2] may be summarized as follows. A series of n predictions $\{\bar{X}_i ; i = 1, \cdots, n\}$ is considered; a prediction $\bar{X}_i$ is a random vector in k-dimensional Euclidean space R^k and we let $\bar{X}_i = \{X_{\mu i} ; \mu = 1, \cdots, k\}$. The distribution of the n predictions is assumed to be Gaussian with independence between the n values of any coordinate and the n values of any other. Letting $\{X_{\mu i} ; i = 1, \cdots, n\}$ be Gaussian with mean $\{m_{\mu i} ; i = 1, \cdots, n\}$ and covariance matrix $\| \sigma_{ij}^{(\mu)} \|$, then $\{X_{\mu i} ; i = 1, \cdots, n\}$ and $\{X_{\nu i} ; i = 1, \cdots, n\}$ are assumed independent for $\mu \neq \nu$. A prediction $\bar{X}_i = \bar{X}_i = (x_{1i}, \cdots, x_{ki})$ becomes a successful prediction with probability given by $s_i(\bar{x}_i)$, the success function. In [2] $s_i(\bar{x}_i)$ has the

Received 10/1/51, revised 11/20/52.

following Gaussian form:

$$s_i(\bar{x}_i) = C_i \exp -\tfrac{1}{2} \sum_{\mu\nu} \tau_{(i)}^{\mu\nu} x_{\mu i} x_{\nu i}, \tag{3.1}$$

where $0 \leqq C_i \leqq 1$, $\| \tau_i^{\mu\nu} \|$ is positive definite, and μ, ν range over the set $\{1, \cdots, k\}$. The general distribution obtained in [2] is the distribution of R, the number of successful predictions.

In applying the results obtained in [2] to the Cunningham and Hynd problem, it is found that the success function is not immediately available in terms of the prediction $\bar{x}$; rather, it is given in terms of a vector $\bar{y}$ which has a Gaussian distribution about $\bar{x}$. The following lemma shows that, if the success function is Gaussian of form (3.1) in terms of $\bar{y}$, then it is also Gaussian in terms of $\bar{x}$.

LEMMA. *If $(Y_1, \cdots, Y_k)$ has a Gaussian distribution with mean $(x_1, \cdots, x_k)$ and covariance matrix $\| G_{\mu\nu} \| = \| G^{\mu\nu} \|^{-1}$, and if the success function in terms of $(y_1, \cdots, y_k)$ is $B \exp -\frac{1}{2}\sum_{\mu\nu} T^{\mu\nu} y_\mu y_\nu$, then the success function in terms of $(x_1, \cdots, x_k)$ is also Gaussian diffuse and has the form $C \exp -\frac{1}{2}\sum_{\mu\nu} \tau^{\mu\nu} x_\mu x_\nu$, where*

$$C = B \mid \delta_{\mu\nu} \sum_{\mu'} G^{\mu\mu'} T^{\mu'\nu} \mid^{-\frac{1}{2}},$$

$$\tau^{\mu\nu} = G^{\mu\nu} - \sum_{\mu'\nu'} \| G^{\mu'\nu'} + T^{\mu'\nu'} \|^{-1} G^{\mu'\mu} G^{\nu'\nu}.$$

PROOF. We calculate the probability of a "hit" as a function of the point of aim $(x_1, \cdots, x_k)$.

$$\begin{aligned}
s(\bar{x}) &= E\{B \exp -\tfrac{1}{2} \sum_{\mu\nu} T^{\mu\nu} Y_\mu Y_\nu\} \\
&= B \frac{\mid G^{\mu\nu} \mid^{\frac{1}{2}}}{(2\pi)^{k/2}} \int \exp\{-\tfrac{1}{2} \sum_{\mu\nu} G^{\mu\nu}(y_\mu - x_\mu)(y_\nu - x_\nu) - \tfrac{1}{2} \sum_{\mu\nu} T^{\mu\nu} y_\mu y_\nu\} \prod_\mu dy_\mu \\
&= B \frac{\mid G^{\mu\nu} \mid^{\frac{1}{2}}}{(2\pi)^{k/2}} \exp\{-\tfrac{1}{2} \sum_{\mu\nu} G^{\mu\nu} x_\mu x_\nu\} \\
&\qquad \cdot \int \exp\{-\tfrac{1}{2} \sum_{\mu\nu} (G^{\mu\nu} + T^{\mu\nu}) y_\mu y_\nu + \sum_\mu [\sum_{\mu'} G^{\mu\mu'} x_{\mu'}] y_\mu\} \prod_\mu dy_\mu \\
&= B \frac{\mid G^{\mu\nu} \mid^{\frac{1}{2}}}{\mid G^{\mu\nu} + T^{\mu\nu} \mid^{\frac{1}{2}}} \exp -\tfrac{1}{2}[\sum_{\mu\nu} G^{\mu\nu} x_\mu x_\nu - \sum_{\mu\nu\mu'\nu'} \| G^{\mu'\nu'} + T^{\mu'\nu'} \|^{-1} G^{\mu'\mu} G^{\nu'\nu} x_\mu x_\nu] \\
&= B \mid \delta_{\mu\nu} + \sum_{\mu'} G_{\mu\mu'} T^{\mu'\nu} \mid^{-\frac{1}{2}} \\
&\qquad \cdot \exp\{-\tfrac{1}{2} \sum_{\mu\nu} [G^{\mu\nu} - \sum_{\mu'\nu'} \| G^{\mu'\nu'} + T^{\mu'\nu'} \|^{-1} G^{\mu'\mu} G^{\nu'\nu}] x_\mu x_\nu\}.
\end{aligned}$$

This completes the proof.

4. The Cunningham and Hynd problem. Cunningham and Hynd were interested in the probability P of at least one hit in a series of n rounds. Using

(3.2) in [2] we have the following expression for P:

$$P = E_1 - E_2 + \cdots + (-1)^{n+1}E_n, \tag{4.1}$$

where

$$E_r = \sum_{\beta_r} E_{\beta_r} \tag{4.2}$$

for which the summation is taken over all sets β_r of r integers chosen from the first n integers and E_{β_r} is the probability that all the rounds designated by the elements of β_r will be hits. Our problem is thus to calculate E_r or E_{β_r} and the above formulas will give the desired probability P. Any other probabilities for the distribution of R will be obtained from formula (3.3) in [2].

We now consider the two-dimensional problem [1] and introduce the following notation and formulation.

The target is given by the probability of a hit on the ith round in terms of the trajectory coordinates (with the center of the target as origin).

$$Pr\{\text{Hit} \mid (y_{1i}, y_{2i})\} = \exp\left\{-\tfrac{1}{2}\frac{y_{1i}^2 + y_{2i}^2}{\sigma_{(i)}^2}\right\},$$

where $\sigma_{(i)}$ is a measure of the effective radius of the target. In the notation of the lemma in Section 3, we have

$$B = 1,$$

$$\| T^{\mu\nu} \| = \begin{Vmatrix} \sigma_{(i)}^{-2} & 0 \\ 0 & \sigma_{(i)}^{-2} \end{Vmatrix}.$$

The dispersion error is given by the covariance matrix on the ith round (independence of coordinates being assumed as in [1]).

$$\| G_{\mu\nu}^{(i)} \| = \begin{Vmatrix} \sigma_{1i}^{2(g)} & 0 \\ 0 & \sigma_{2i}^{2(g)} \end{Vmatrix},$$

where $\sigma_{\mu i}^{2(g)}$ is the variance of the μth coordinate of the dispersion error.

The success function is given, using the lemma in Section 3, by the following:

$$C_i = \frac{\sigma^{2(i)}}{\sigma_{1i}^{(i+g)}\,\sigma_{2i}^{(i+g)}},$$

$$\| \tau_{(i)}^{\mu\nu} \| = \begin{Vmatrix} \sigma_{1i}^{-2(i+g)} & 0 \\ 0 & \sigma_{2i}^{-2(i+g)} \end{Vmatrix},$$

where $\sigma_{\mu i}^{2(i+g)} = \sigma^{2(i)} + \sigma_{\mu i}^{2(g)}$, the addition of variances.

The aiming error is given by the variance of that error for μth coordinate on the ith round, $\sigma_{ii}^{(\mu)} = \sigma_{\mu i}^{2(a)}$, and the correlation between the values of the μth coordinate for rounds i and j, $\rho_{ij}^{(\mu)}$. We are assuming that the mean is equal to zero; that is, there is no bias in the aiming.

Using Theorem 3 in [2] we obtain the following expression for E_{β_r}:

$$E_{\beta_r} = \prod_{p\varepsilon\beta_r} C_p \prod_{\mu=1,2} | \delta_{pq} + \sigma_{pq}^{(\mu)} \tau_{(p)}^{\mu\mu} |^{-\frac{1}{2}}$$

$$= \prod_{p\varepsilon\beta_r} \frac{\sigma^{2(t)}}{\sigma_{1p}^{(t+g)} \sigma_{2p}^{(t+g)}} \prod_{\mu} | \delta_{pq} + \sigma_{\mu p}^{(a)} \sigma_{\mu q}^{(a)} \rho_{pq}^{(\mu)} \sigma_{\mu p}^{-2(t+g)} |^{-\frac{1}{2}} \tag{4.3}$$

$$= \prod_{p\varepsilon\beta_r} \frac{\sigma^{2(t)}}{\sigma_{1p}^{(a)} \sigma_{2p}^{(a)}} \prod_{\mu=1,2} \left| \rho_{pq}^{(\mu)} + \delta_{pq} \frac{\sigma_{\mu p}^{2(t+g)}}{\sigma_{\mu p}^{2(a)}} \right|^{-\frac{1}{2}}.$$

Considering now the application of the Type I limiting distribution, we have the following result:

$$E_r = \frac{1}{r!\,T^r} \int_0^T \cdots \int_0^T \prod_{p=1}^{r} \frac{n\sigma^{2(t)}}{\sigma_{1(t_p)}^{(a)} \sigma_{2(t_p)}^{(a)}} \prod_{\mu} \left| \rho_{(t_p t_q)}^{(\mu)} + \delta_{pq} \frac{\sigma_{\mu(t_p)}^{2(t+g)}}{\sigma_{\mu(t_p)}^{2(a)}} \right|^{-\frac{1}{2}} \prod_{1}^{r} dt_p . \tag{4.4}$$

The duration of the burst is T, and $\sigma_{1(t_p)}^{2(a)}$, for example, is the variance of the aiming error at time t_p. From the conditions of Theorem 4 in [2], (4.4) will be a valid approximation for (4.2) with (4.3) if $\sigma^{2(t)}$ is of order $1/n$ with respect to $\sigma^{2(g)}$ and n is large.

Similarly the Type II limiting distribution gives the following:

$$E_r = \frac{1}{r!\,T^r} \int_0^T \cdots \int_0^T \prod_{p=1}^{r} \frac{n\sigma^{2(t)}}{\sigma_{1(t_p)}^{(a)} \sigma_{2(t_p)}^{(a)}} \prod_{\mu=1,2} \left| \rho_{(t_p t_q)}^{(\mu)} \right|^{-\frac{1}{2}} \prod_{p=1}^{r} dt_p . \tag{4.5}$$

The limiting conditions are that $\sigma_{1i}^{2(t+g)}$, $\sigma_{2i}^{2(t+g)}$ be of order $1/n$,

$$\frac{\sigma^{2(t)}}{\sigma_{1i}^{(t+g)} \sigma_{2i}^{(t+g)}}, \qquad \sigma_{1i}^{2(a)}, \text{ and } \sigma_{2i}^{2(a)} \text{ of order 1, and } n \text{ large.}$$

5. Evaluation for $n = 5$. The evaluation of the probability P requires all values of E_r or a sufficient number to estimate P from a truncated portion of the series (4.1). This direct evaluation can readily be carried out if n is small as in the example ($n = 5$) in this section; we defer to Section 6 a consideration of the evaluation for large n.

We use the data in [1] for a one-second burst of $n = 5$ rounds. As indicated in [1] we take the correlation $\rho_{ij}^{(\mu)}$ to be the same for the horizontal and vertical coordinates and to be dependent only on the time interval between the rounds. Then from the experimental correlations tabulated in [1] we have $\rho(0) = 1.00$, $\rho(.25) = .80$, $\rho(.50) = .62$, $\rho(.75) = .48$, $\rho(1.00) = .33$. From formulas (4.1), (4.2), and (4.3) we have that

$$P = \beta \frac{n}{\alpha} - \beta^2 \sum_{i_1<i_2} \frac{1}{\alpha^2 - \rho^2(i_2 - i_1)} + \cdots - (-\beta)^r \sum_{i_1<\cdots<i_r} | \rho(i_p - i_q)(1 - \delta_{pq}) + \alpha\delta_{pq} |^{-1}, \tag{5.1}$$

where

$$\beta = \frac{\sigma^{2(t)}}{\sigma^{2(a)}},$$

$$\alpha = 1 + \frac{\sigma^{2(t+g)}}{\sigma^{2(a)}},$$

and the variances $\sigma^{2(t)}$, $\sigma^{2(g)}$, $\sigma^{2(a)}$ are assumed to remain constant for the duration of the burst. The function P was calculated for a series of values of α and β and plotted in Fig. 1 with $\sigma^{(t)} = 1$.

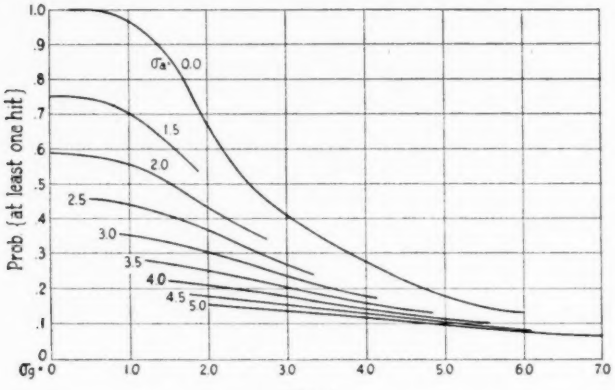

FIG. 1

6. Evaluation for large n. The direct computation of P for a series of large values of n would be excessive. We now introduce a procedure for approximation. Assuming that the correlation depends only on the time interval between the rounds, that the horizontal and vertical components have the same distribution, and that variances are constant over the time interval, then P has the form (5.1).

Consider for a moment the case in which correlation is absent ($\rho_{ij} = \delta_{ij}$), and let

$$F_r = E_r(\rho_{ij} = \delta_{ij}) = \binom{n}{r}\left[\frac{\beta}{\alpha}\right]^r.$$

The expression for P also simplifies:

$$P = \beta\frac{n}{\alpha} - \beta^2\binom{n}{2}\frac{1}{\alpha^2} + \cdots - (-\beta)^r\binom{n}{r}\frac{1}{\alpha} + \cdots.$$

This binomial expression has simple terms and it is natural to investigate the correction factor between terms in the correlated and uncorrelated series. We

therefore define as follows:

$$E_r = F_r(1 + c_r), \qquad c_r \geqq 0, \qquad c_1 \equiv 0.$$

We now derive an expression for c_r from which it easily follows that c_r is non-negative.

$$c_r = \alpha^r \sum_{i_1 < \cdots < i_r} \frac{|\, \alpha\delta_{pq} + \rho_{pq}(1 - \delta_{pq})\,|^{-1}}{\binom{n}{r}} - 1$$

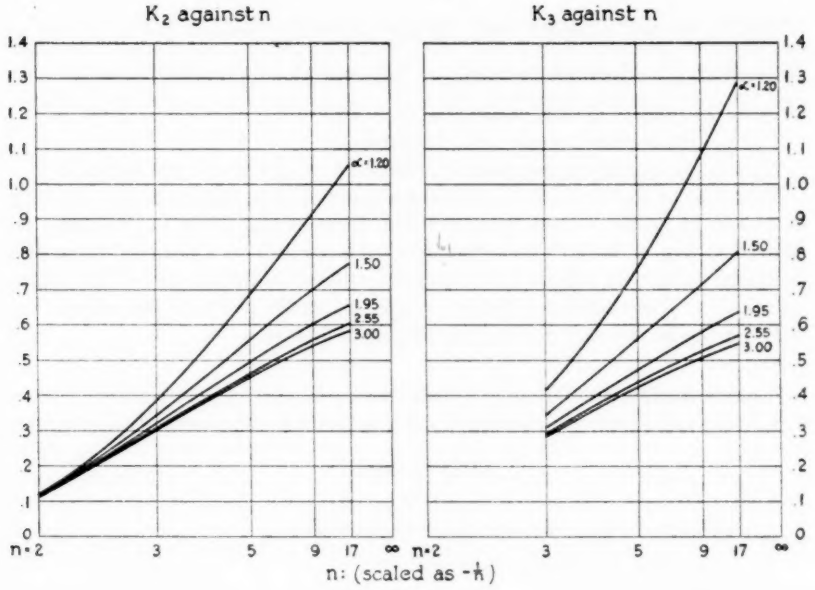

FIG. 2

$$\frac{1}{\binom{n}{r}} \sum_{\beta_r} \left[\frac{1}{\delta_{pq} + \dfrac{\rho_{pq}}{\alpha}(1 - \delta_{pq})} - 1 \right].$$

From Theorem 4 in [2] we know that c_r approaches a limit as $n \to \infty$. This can also be seen directly since $\alpha > 1$ ($\alpha = 1$ would imply there was no target!). The limiting value is derived from Theorem 4, as

$$\lim_{n \to \infty} C_r = \frac{1}{T^r} \int_0^T \cdots \int_0^T \left| \delta_{pq} + \frac{\rho(t_p - t_q)}{\alpha}(1 - \delta_{pq}) \right|^{-1} \alpha t_1 \cdots \alpha t_r - 1.$$

Since c_r is stable for large n we now investigate the dependence on r. We note that $\lim_n c_r$ is the excess over 1 of the average value of the reciprocal of a deter-

minant having 1's down the diagonal, and that these 1's are larger by $(\alpha - 1)/\alpha$ than elements which would be sufficient to make the matrix positive definite.

Thus assuming $\alpha > 1$ and expanding the determinant we find that the determinant is $1 - \sum_{p<q}\{\rho(t_p - t_q)/\alpha\}^2 \cdots$, where p, q range over $(i_1, \cdots, i_r)$. Formally taking the reciprocal, we find that the inverse of the determinant is $1 - \sum_{p<q}\{\rho(t_p - t_q)/\alpha\}^2 \cdots$. The excess over 1 of the average of such expressions will have a first term which is the sum of $\binom{r}{2}$ squares of the form

$$[\rho(t_p - t_q)/\alpha]^2.$$

This expression suggests replacing our correction factor c_r by k_r defined by $k_r = \alpha^2 c_r/\binom{r}{2}$. Thus we have $E_r = F_r[1 + \binom{r}{2} k_r/\alpha^2]$. The correction constant k_r has the following properties:

(1) k_r approaches a limit as n increases.

(2) k_r is to the first approximation independent of r and α for α large.

Values of k_2 and k_3 were calculated for a series of n, using the time interval $T = 1$ as in Section 5 and the correlation function $\rho(t)$ given in [1]. These are plotted in Fig. 2. It is to be noted that the approach to a limiting value as n increases seems very regular and k_2 and k_3 are quite similar except for the smaller values of α. This stability of the functions $\{k_r\}$ would facilitate the calculations if P were to be obtained for a series of large values of n.

7. Acknowledgment. The author wishes to express his appreciation to Professor S. S. Wilks and Professor John W. Tukey for valuable discussions of the problem.

REFERENCES

[1] L. B. C. Cunningham and W. R. B. Hynd, "Random processes in air warfare," *J. Roy. Stat. Soc., Suppl.*, Vol. 8 (1946), pp. 62–85.

[2] D. A. S. Fraser, "Generalized hit probabilities with a Gaussian target," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 248–255.

NOTES

A CLASS OF MINIMAX TESTS FOR ONE-SIDED COMPOSITE HYPOTHESES[1]

By S. G. Allen, Jr.

Stanford University

Summary. Fixed sample-size procedures are considered for testing a one-sided composite hypothesis concerning a real, one-dimensional parameter of an exponential distribution (1.1). In particular, conditions are studied such that the minimax tests have a critical region which is a semi-infinite interval on the real line.

1. Statement of the problem. Let X be a real-valued, one-dimensional random variable with the probability density

$$p(x, \theta) = \omega(\theta)\psi(x)e^{\theta x}, \tag{1.1}$$

where

$$\omega(\theta) = \left[\int_{-\infty}^{\infty} e^{\theta x}\psi(x)\,dx\right]^{-1} \tag{1.2}$$

is a positive, bounded, continuous function of the real variable θ and where $\psi(x)$ is a continuous, nonnegative function of the real variable x. Let $X_1, X_2, \cdots, X_n$ denote n independent observations on X, and let $T(X_1, \cdots, X_n)$ denote a fixed sample-size procedure based on the n observations for testing the composite hypothesis $\theta > \theta_0$ against the alternative $\theta < \theta_0$. The loss functions are defined as follows: if the hypothesis is rejected, the loss is $w_1(\theta) \geqq 0$ for $\theta > \theta_0$ and $w_1(\theta) = 0$ otherwise; if the hypothesis is accepted, the loss is $w_2(\theta) \geqq 0$ for $\theta < \theta_0$ and $w_2(\theta) = 0$ otherwise. Furthermore, it is assumed that the function $w_1(\theta)$ is actually positive for at least one value of $\theta > \theta_0$, and $w_2(\theta)$ is positive for at least one value of $\theta < \theta_0$. The problem to be considered is the selection of a minimax test procedure $T(X_1, \cdots, X_n)$ under these conditions.

2. A class of minimax tests. For testing the simple hypothesis $\theta = \theta_2$ in (1.1)

Received 11/5/51, revised 11/2/52.

[1] This paper, representing work done under the sponsorship of the Office of Naval Research, was presented at the Western meetings of the Institute of Mathematical Statistics, June 15–16, 1951, Santa Monica, California. Discussion with members of the Department of Statistics, Stanford University, in particular with Professor M. A. Girshick, was most beneficial in the formulation of the present draft of the paper. The author understands that results similar to those of the present paper were obtained for the sequential case by Milton Sobel in his doctoral thesis, "An essentially complete class of decision functions for certain standard sequential problems."

against the simple alternative $\theta = \theta_1$ with $\theta_1 < \theta_2$, the minimax procedure based on n independent observations on X is well known [1]. The value of the statistic

$$\lambda = \lambda(\theta_1, \theta_2) = \prod_{i=1}^{n} \frac{p(x_i, \theta_2)}{p(x_i, \theta_1)} \tag{2.1}$$

is computed from the observed values of X in the sample. The hypothesis is then accepted if $\lambda > c$ and rejected if $\lambda \leqq c$, where the criterion c satisfies

$$w_1(\theta_2)Pr(\lambda \leqq c \mid \theta_2) = w_2(\theta_1)Pr(\lambda > c \mid \theta_1). \tag{2.2}$$

This value of c is

$$c = \frac{w_2(\theta_1)g}{w_1(\theta_2)(1 - g)}, \tag{2.3}$$

where g is the least favorable a priori probability that $\theta = \theta_1$.

From the form of the density function (1.1), it is clear that an identical procedure to the preceding ratio test specifies acceptance of the hypothesis if and only if $\sum_{i=1}^{n} x_i > k$, where

$$k = \frac{1}{\theta_2 - \theta_1} \log c \left[\frac{\omega(\theta_1)}{\omega(\theta_2)}\right]^n. \tag{2.4}$$

Since the probability density of the statistic $\sum_{i=1}^{n} x_i$ is again of the form (1.1) (see Section 4 of [2]), the discussion of tests like the above is not restricted by an assumption that the sample consists of a single observation on X. Therefore, the number k defined in (2.4) may be determined by a condition equivalent to (2.2) with $n = 1$, namely,

$$w_1(\theta_2)Pr(X \leqq k \mid \theta_2) = w_2(\theta_1)Pr(X > k \mid \theta_1). \tag{2.5}$$

Let $T_k(X)$ denote a test procedure specifying acceptance of the hypothesis $\theta > \theta_0$ if the observed value of X exceeds k, and specifying rejection otherwise. One might ask if such test procedures, which form a class of minimax procedures in the case of the simple dichotomy, retain this property in the more general problem of Section 1. If so, does a condition similar to (2.5) determine the minimax test?

The following theorem supplies an answer.[2]

THEOREM 1. *Let*

$$R_1(k, \theta) = w_1(\theta) \int_{-\infty}^{k} \omega(\theta)\psi(x)e^{\theta x}\, dx, \qquad \theta \geqq \theta_0, \tag{2.6}$$

$$R_2(k, \theta) = w_2(\theta) \int_{k}^{\infty} \omega(\theta)\psi(x)e^{\theta x}\, dx, \qquad \theta \leqq \theta_0. \tag{2.7}$$

[2] The motivating idea for Theorem 1 was a lot acceptance sampling procedure proposed in an unpublished paper by Mr. Norman Rudy of Sacramento State College.

Then $T_k(X)$ is minimax if

$$\max_{\theta \geq \theta_0} R_1(k, \theta) = \max_{\theta \leq \theta_0} R_2(k, \theta). \tag{2.8}$$

PROOF. Let $R(T, G)$ denote the expected loss of a test T with respect to the a priori distribution with cdf $G(\theta)$. In particular, for a k_0 satisfying (2.8),

$$\begin{aligned} R(T_{k_0}, G) &= \int_{\theta_0}^{\infty} R_1(k_0, \theta)\, dG(\theta) + \int_{-\infty}^{\theta_0} R_2(k_0, \theta)\, dG(\theta) \\ &\leqq \int_{\theta_0}^{\infty} \max_{\theta \geq \theta_0} R_1(k_0, \theta)\, dG(\theta) + \int_{-\infty}^{\theta_0} \max_{\theta \leq \theta_0} R_2(k_0, \theta)\, dG(\theta) \qquad (2.9) \\ &= \max_{\theta \geq \theta_0} R_1(k_0, \theta) = \max_{\theta \leq \theta_0} R_2(k_0, \theta). \end{aligned}$$

Let θ_1, θ_2 be values of θ such that $\theta_1 \leqq \theta_0 \leqq \theta_2$ and

$$\max_{\theta \geq \theta_0} R_1(k_0, \theta) = R_1(k_0, \theta_2), \tag{2.10}$$

$$\max_{\theta \leq \theta_0} R_2(k_0, \theta) = R_2(k_0, \theta_1). \tag{2.11}$$

If G is a distribution concentrating all probability at θ_1 and θ_2, then the equality sign holds throughout (2.9). Therefore

$$\max_{G} R(T_{k_0}, G) = \max_{\theta \geq \theta_0} R_1(k_0, \theta) = \max_{\theta \leq \theta_0} R_2(k_0, \theta). \tag{2.12}$$

In particular let G_0 be the distribution given by $g = Pr(\theta = \theta_1)$, $1 - g = Pr(\theta = \theta_2)$, where g satisfies

$$k_0 = \frac{1}{\theta_2 - \theta_1} \log \frac{w_2(\theta_1) g \omega(\theta_1)}{w_1(\theta_2)(1 - g)\omega(\theta_2)}.$$

Clearly T_{k_0} is the Bayes procedure against G_0. (Compare g in (2.3) and (2.4).) Hence

$$\min_{T} R(T, G_0) = R(T_{k_0}, G_0) = \max_{G} R(T_{k_0}, G).$$

Application of the saddle-point theorem of [3] completes the proof.

3. An example based on the normal distribution. Suppose it is desired to test the hypothesis that θ, the mean of a normal distribution with variance one, is positive against the alternative that it is negative, where $w_1(\theta) = \theta$ for $\theta \geqq 0$ and $w_2(\theta) = -\theta$ for $\theta \leqq 0$. The functions defined in (2.6) and (2.7) are

$$R_1(k, \theta) = \int_{-\infty}^{k-\theta} \frac{\theta}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy, \qquad \theta \geqq 0$$

$$R_2(k, \theta) = \int_{k-\theta}^{\infty} -\frac{\theta}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy = \int_{-\infty}^{-k+\theta} -\frac{\theta}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy, \qquad \theta \leqq 0.$$

Since $R_2(k, -|\theta|) \equiv R_1(-k, |\theta|)$, it follows that $\max_{\theta \leq 0} R_2(0, \theta) = \max_{\theta \geq 0} R_1(0, \theta)$, provided the latter exist. This is certainly the case, since, by L'Hospital's rule,

$$\lim_{\theta \to \infty} R_1(0, \theta) = \lim_{\theta \to \infty} \frac{\theta^2}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2} = 0.$$

4. Remarks on the discrete case. The continuous distributions studied in the preceding sections represent a sub-family of a more general family of distributions of the form $\omega(\theta)e^{\theta x}d\Psi(x)$, where Ψ is a measure on the real numbers and where

$$\omega(\theta) = \left[\int_{-\infty}^{\infty} e^{\theta x}\, d\Psi(x)\right]^{-1}$$

is a positive bounded function of the real variable θ. This family includes many of the most important distributions encountered in statistics, such as the normal, χ^2, binomial, negative binomial, and Poisson distributions.

Suppose the distribution under consideration in this family is a discrete one, and suppose that $\Psi(x)$ assumes jumps at each value of a denumerable, ordered sequence $(x_1, x_2, \cdots)$. For example, if X is the number of successes in n Bernoulli trials, the function $\Psi(x)$ assumes jumps at $x = 0, 1, 2, \cdots, n$. In general, it will not be possible to find a value of k in such a sequence so that condition (2.8) is fulfilled. However, a randomized mixture of two procedures T_k and $T_{k'}$ will be a minimax procedure if there exists a pair (k, k') such that

$$\max_{\theta \geq \theta_0} R_1(k', \theta) < \max_{\theta \leq \theta_0} R_2(k', \theta),$$

$$\max_{\theta \geq \theta_0} R_1(k, \theta) > \max_{\theta \leq \theta_0} R_2(k, \theta),$$

where k' is the next smaller element than k in the sequence $(x_1, x_2, \cdots)$. In this event, the minimax test procedure consists of the following: reject the hypothesis $\theta > \theta_0$ if $x < k$; accept the hypothesis if $x > k$; if $x = k$, accept the hypothesis with probability f and reject with probability $1 - f$, where f satisfies

$$\max_{\theta \geq \theta_0} [fR_1(k', \theta) + (1 - f)R_1(k, \theta)] = \max_{\theta \leq \theta_0} [fR_2(k', \theta) + (1 - f)R_2(k, \theta)].$$

REFERENCES

[1] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, 1950, p. 17.

[2] M. A. Girshick and L. J. Savage, "Bayes and Minimax Estimates for Quadratic Loss Functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53–73.

[3] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, 2nd ed., Princeton University Press, 1947, p. 95.

NOTE ON COMPUTATION OF ORTHOGONAL PREDICTORS

By Bradford F. Kimball

New York State Department of Public Service

Summary. The present note calls attention to a simple algorithm for computation of the orthogonal matrix associated with the matrix of the normal equations of least squares. An application of the "forward" solution associated with the orthogonalization process is also pointed out.

1. Introduction. This note is supplementary to Dwyer's recent book *Linear Computations*, [1]. The author also acknowledges reference to extensive unpublished notes on matrix analysis of numerical methods kindly furnished him by A. S. Householder.

Orthogonalization of the predicting variables of the least square problem has been considered since the introduction of orthogonal polynomials into the least-squares problem by Tchebycheff 1853–73 and the significant doctoral dissertation of the Danish mathematician and actuary, J. P. Gram, in 1879, [3]. A discussion of the problem is also to be found in Poincaré's *Calcul des Probabilités*, 1912, Chapter XV. Various methods of simplifying the computations for purposes of mass application to statistical data have been studied since then. A complete bibliography is beyond the scope of this note.

The advantages of obtaining the solution of the least squares problem in terms of orthogonal predictors are numerous. Perhaps the most obvious are those associated with the resulting simplified expressions for the error formulae, correlations [2] and sampling variance of the fitted function [5]. Also the orthogonalization of the predicting variables is a starting point for the computation of principal components significant to structural relationships in psychology and economics. A further application of the associated "forward" solution is pointed out in connection with the lemma stated in Section 4 of the present paper.

Two solutions are obtained. One is in terms of a slight extension of the algorithm of the square root method of solving the normal equations [2], and the other is in terms of the algorithm of the Gauss-Doolittle method. The first method has the advantage that the coefficients of the orthogonal predictors have unit sampling variance weight. The second method is based on the more familiar Doolittle algorithm and does not involve square roots.

There seems to be some need for standardization of notation in matrix analysis. Since Householder has presented a consistent application of matrix analysis to a large body of material, we shall keep to his notation in the following respects. Capitals will be used for matrices and the transpose of a matrix will be indicated by the superscript T rather than by a prime. Single row or column matrices and vectors will be denoted by small letters.

Received 3/17/52.

2. Derivation of orthogonal matrix from matrix theory. Let

$$X = \| x_{1j}, x_{2j}, \cdots, x_{nj} \|, \qquad j = 1, 2, \cdots N, n \leqq N \tag{2.1}$$

denote the N rowed matrix of the N observations on the n predictor variables x_{ij}. With the subscript j suppressed, the observations on the n original predictor variables are treated as column vectors x_i, $i = 1$ to n.

Consider a matrix V of n rows and columns. The product matrix XV^T will consist of n columns, each of which is a linear transform of the n columns of the data matrix X. Hence, if the n linear functions defined in the n columns are to be orthonormal, the following matrix relation must obtain

$$(XV^T)^T(XV^T) = I, \tag{2.2}$$

where I denotes the identity matrix.

From general matrix theory $(XV^T)^T = VX^T$, and (2.2) can be written [4]

$$V(X^TX)V^T = I. \tag{2.3}$$

For purposes of setting up an algorithm for computation of V we introduce the triangular matrices of the Gauss-Doolittle and square root methods. Using Dwyer's notation, but replacing small letters by capitals where matrices are referred to, the upper triangular matrix of the summation rows of the Doolittle process is denoted by G^T, and the corresponding matrix (after division of the kth row by the square root of g_{kk}) for the square root method is denoted by S (see [1] p. 188). Let D denote the diagonal matrix composed of the diagonal elements of G. Then

$$S = D^{-\frac{1}{2}}G^T. \tag{2.4}$$

Denoting the matrix of the moments of the predictor variables by A, the fact that S factors A is expressed as

$$A = X^TX = S^TS. \tag{2.5}$$

Substituting in (2.3) we find $V(S^TS)V^T = I$ which can be written

$$(SV^T)^T(SV^T) = I. \tag{2.6}$$

A meaningful solution of this equation is given by

$$SV^T = I, \qquad V^T = S^{-1}, \qquad V = (S^T)^{-1} \tag{2.7}$$

and using (2.4) the solution can also be written as

$$V^T = (G^T)^{-1}D^{\frac{1}{2}}. \tag{2.8}$$

3. Computational algorithm for orthogonal multipliers. Dwyer has pointed out that the inverse of the matrix S can be very directly computed by using the identity matrix for the reference matrix of the dependent variable on the right in a computational schedule similar to the familiar Doolittle algorithm ([1] pp. 191 and 197, explicit in Table 13.6b and implicit in Table 13.8a).

For reference purposes we reproduce the schedule for $n = 3$ in terms of symbols, omitting the secondary subscripts:

$$A = \begin{matrix} a_{11} & a_{12} & a_{13} \\ * & a_{22} & a_{23} \\ * & * & a_{33} \end{matrix} \qquad \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} = I$$

$$S = \begin{matrix} s_{11} & s_{12} & s_{13} \\ 0 & s_{22} & s_{23} \\ 0 & 0 & s_{33} \end{matrix} \qquad \begin{matrix} v_{11} & 0 & 0 \\ v_{21} & v_{22} & 0 \\ v_{31} & v_{32} & v_{33} \end{matrix} = V.$$

Because of the symmetry of the A matrix the computational algorithm assures that $SV^T = I$. Denote the n columns of the product matrix XV^T by the column vectors

$$\phi_1 = v_{11}x_1\,, \qquad \phi_2 = v_{21}x_1 + v_{22}x_2\,, \quad \cdots \phi_n = \sum v_{ni}x_i\,. \tag{3.1}$$

These vectors constitute a set of n mutually orthogonal vectors which are linear transforms of the original set. They take the classical form, which one would expect, and are a "normal" set, since $\phi_k^T\phi_k = 1$.

For the uninitiated who are not familiar with the square-root method, or who prefer to use the Gauss-Doolittle method in the solution of the problem of least squares, recall from (2.8) that $V = D^{\frac{1}{2}}G^{-1}$ and note that G^{-1} is given by the triangular matrix represented by the lower of the doublet of rows extending under the identity reference matrix of the Doolittle algorithm ([1] p. 191, Table 13.6a). It follows that the triangular matrix from the upper of the doublet of rows, which we denote by R, satisfies the relation

$$R = D^{\frac{1}{2}}V. \tag{3.2}$$

Thus a set of mutually orthogonal vectors ψ_k is determined by the columns of the matrix product XR^T. These, although simpler to compute than the ϕ_k, have the disadvantage that $\psi_k^T\psi_k = g_{kk}$. One can of course compute ϕ_k from ψ_k by simply dividing through by $\sqrt{g_{kk}}$.

4. Accessory relations. Let y denote a column vector of N observations on the variable to be fitted in a least squares problem and let t denote the column vector of the coefficients of the orthogonal predictors ϕ_k. Recalling that $\Phi = XV^T$ represents an n column, N row matrix, conventional solution of the normal equations by matrix analysis leads to [4]

$$t = \Phi^T y, \qquad t_k = \phi_k^T y. \tag{4.1}$$

Furthermore it is easily seen from matrix analysis that

$$\Phi^T X = VX^T X = VS^T S = IS = S$$

and hence

$$s_{ki} = \phi_k^T x_i\,, \qquad i \neq 0. \tag{4.2}$$

Since the algorithm for finding s_{j0} in the y column is operationally the same as that used for finding s_{ik} in the x_k column, it follows that

$$s_{k0} = \phi_k^T y = t_k \tag{4.3}$$

and hence the fitted function u is given explicitly by

$$\begin{aligned} u &= \sum t_k \phi_k = \sum s_{k0} \phi_k , \\ \phi_k &= v_{k1} x_1 + v_{k2} x_2 + , \cdots , + v_{kk} x_k , \qquad k = 1, 2, \cdots n. \end{aligned} \tag{4.4}$$

Clearly if computations are based on the Doolittle algorithm and the predictors ψ_k are used, similar relations will hold:

$$g_{ki} = \psi_k^T x_i , \qquad i \neq 0; \; g_{k0} = \psi_k^T y, \tag{4.5}$$

and the coefficients of the predicting variables ψ_k will be given by

$$b_{k0} = g_{k0}/g_{kk} . \tag{4.6}$$

A forward solution is furnished by solving the above relations for the coefficients of x_i in the fitted function. Denoting these coefficients by the column vector h, we have $u = \Phi t = Xh$. Since X is not in general a square matrix, this equation is solved for h by multiplying it by what has been called the "pseudo-inverse" of X; namely, $(X^TX)^{-1}X^T$. The result is

$$h = (X^TX)^{-1}(X^T\Phi t) = (X^TX)^{-1}(X^TX)V^Tt = V^Tt.$$

In explicit, nonmatrix form

$$\begin{aligned} h_1 &= t_1 v_{11} + t_2 v_{21} + \cdots + t_n v_{n1} \\ h_2 &= t_2 v_{22} + \cdots + t_n v_{n2} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ h_{n-1} &= t_{n-1} v_{n-1,n-1} + t_n v_{n,n-1} \\ h_n &= t_n v_{nn} . \end{aligned} \tag{4.7}$$

A useful point about this forward solution which may easily be passed over without recognition is the following.

LEMMA. *If* t_1, t_2 $\cdots$ t_k *are determined as* $\phi_1^T y$, $\phi_2^T y$, $\cdots$, $\phi_k^T y$, *then* $h_1, h_2, \cdots, h_k$ *derived from the above schedule* (4.7) *will satisfy the first* k *normal equations of the* n *predictor problem for arbitrary values of* $t_{k+1}, t_{k+2}, \cdots, t_n$.

The proof follows by applying the first k conditions for minimizing the sum of the squares of the deviations to the orthogonal form of the solution (4.4).

The writer has found application of this lemma to the following problem. Several linear functions are to be fitted to separate groups of data by least squares, where, say, one of the coefficients is to be determined so that it is optimum for *all* the groups *lumped together*, and the other coefficients are to be determined separately for *each separate group*. (One such application will be discussed by the author in an article which will appear shortly in the *Journal of the American Statistical Association*.)

REFERENCES

[1] P. S. DWYER, *Linear Computations*, John Wiley and Sons, 1951.

[2] P. S. DWYER, "The square root method and its use in correlation and regression," *J. Amer. Stat. Assn.*, Vol. 40 (1945), pp. 493–503.

[3] J. P. GRAM, "Ueber die entwicklung reeler functionen in reihen," Crelle's *J. für Math.*, Vol. 94 (1883), pp. 41–73. (This is a presentation in German of his 1879 dissertation.)

[4] A. S. HOUSEHOLDER, "Some numerical methods for solving systems of linear equations," *Amer. Math. Monthly*, Vol. 57 (1950), pp. 453–459.

[5] B. F. KIMBALL, "Orthogonal polynomials applied to least square fitting of weighted observations," *Ann. Math. Stat.*, Vol. 11 (1940), p. 352.

ON THE ASYMPTOTIC NORMALITY OF CERTAIN RANK ORDER STATISTICS[1]

BY MEYER DWASS

Northwestern University

1. Summary. Let $(R_1, \cdots, R_N)$ be a random vector which takes on each of the $N!$ permutations of the numbers $(1, \cdots, N)$ with equal probability, $1/N!$. Sufficient conditions are given for the asymptotic normality of $S_N = \sum_{i=1}^{N} a_{Ni} b_{NR_i}$, where $(a_{N1}, \cdots, a_{NN})$, $(b_{N1}, \cdots, b_{NN})$ are two sets of real numbers given for every N. These sufficient conditions are apparently quite different from those given by Wald and Wolfowitz [9] and extended by various writers [4, 7]. In some situations the conditions given here may be easier to apply than those given previously. The most general conditions available to date appear to be those of Hoeffding [4]. In the examples below, however, is given a case of an S_N which does not satisfy the conditions required by Hoeffding's theorem but which is asymptotically normal by our results.

2. Statement of theorem and its proof. We will assume hereafter that

$$\sum_{i=1}^{N} a_{Ni} = \sum_{i=1}^{N} b_{Ni} = 0, \qquad \sum_{i=1}^{N} a_{Ni}^2 = 1.$$

THEOREM. Suppose for an integer $k \geqq 1$ there is a random variable X satisfying the following conditions:

(a) X has a continuous cdf $F(x)$,

(b) if $X_1, \cdots, X_N$ are independent random variables each with the cdf $F(x)$ and $Z_{N1} \leqq \cdots \leqq Z_{NN}$ are the ordered values of $X_1, \cdots, X_N$ then

$$b_{Ni} = EZ_{Ni}^k - \sum_{j=1}^{N} EZ_{Nj}^k/N$$

for all N and i.

Received 7/21/52, revised 11/24/52.

[1] Work done under the sponsorship of the Office of Naval Research while the author was a graduate student at the University of North Carolina.

(c) $E \mid X \mid^{2k} < \infty$.

(d) Either X_i^k is normal or (e) $\max_{1 \leq i \leq N} \mid a_{Ni} \mid \to 0$ as $N \to \infty$.

Then S_N is asymptotically normally distributed.

PROOF OF THEOREM. Associate with the random vector $X_1, \cdots, X_N$ the random vector $R_1, \cdots, R_N$ where R_i = number of $X_j \leqq X_i$.

Let $g_N(\boldsymbol{X}) = g_N(X_1, \cdots, X_N)$ be the random variable $g_N(\boldsymbol{X}) = \sum_{i=1}^N a_{Ni} b_{NR_i}$. Hence, for every N, the distribution of $g_N(\boldsymbol{X})$ is identical with that of S_N, for each assumes the same set of values with the same probabilities. Write

$$g_N(\boldsymbol{X}) = \sum_{i=1}^N a_{Ni} X_i^k - \left(\sum_{i=1}^N a_{Ni} X_i^k - g_N(\boldsymbol{X}) \right).$$

If it can be shown that

$$\sum_{i=1}^N a_{Ni} X_i^k - g_N(\boldsymbol{X}) \tag{1}$$

converges in probability to zero, then if $\sum_{i=1}^N a_{Ni} X_i^k$ has a limiting distribution, $g_N(\boldsymbol{X})$ will approach that same limiting distribution (as $N \to \infty$) ([1], p. 254).

That $\sum_{i=1}^N a_{Ni} X_i^k$ has a limiting normal (0, 1) distribution is seen by applying the condition of Liapounoff that

$$\frac{\left(\sum_{i=1}^N \mid a_{Ni} \mid^3 E \mid X^k - EX^k \mid^3 \right)^{\frac{1}{3}}}{(E(X^k - EX^k)^2)^{\frac{1}{2}}} \to 0$$

as $N \to \infty$. This is so, since

$$\sum_{i=1}^N \mid a_{Ni} \mid^3 \leq \max_{1 \leq i \leq N} \mid a_{Ni} \mid \sum_{j=1}^N (a_{Nj})^2 = \max_{1 \leq i \leq N} \mid a_{Ni} \mid .$$

To show that (1) converges in probability to zero, it will be sufficient to show that $\lim_{N \to \infty} E(\sum_{i=1}^N a_{Ni} X_i^k - g_N(\boldsymbol{X}))^2 = 0$. Denote by U_N the expression

$$U_N = E \left(\sum_{i=1}^N a_{Ni} X_i^k - g_N(\boldsymbol{X}) \right)^2 = E \left(\sum_{i=1}^N a_{Ni} (X_i^k - EX_i^k) - g_N(\boldsymbol{X}) \right)^2$$

$$= E(X^k - EX^k)^2 - \frac{2}{N!} \sum{}' \left[\left(\int N! \sum_{i=1}^N a_{Ni} (X_i^k - EX_i^k) \prod_{i=1}^N dF(x_i) \right) \cdot \sum a_{Ni} b_{Nr_i} \right] + E g_N^2(\boldsymbol{X})$$

where the integral is over that part of the space where $R_i = r_i$ $(i = 1, \cdots, N)$ and $r_1, \cdots, r_N$ is one of the $N!$ permutations of $1, \cdots, N$ and where the summation $\sum'$ is over all such permutations.

By condition (b) and by the fact that $N^{-1} \sum_{i=1}^N EZ_{Ni}^k = EX^k$, it follows that $U_N = E(X^k - EX^k)^2 - E g_N^2(\boldsymbol{X})$. By straightforward algebra,

$$E g_N^2(\boldsymbol{X}) = \frac{1}{N!} \sum{}' \left(\sum_{i=1}^N a_{Ni} b_{Nr_i} \right)^2 = \frac{1}{N-1} \sum_{i=1}^N b_{Ni}^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}(EZ_{Ni}^{k})^{2} - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} EZ_{Ni}^{k}\right)^{2}$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}(EZ_{Ni}^{k})^{2} - \frac{N}{N-1}(EX_{k})^{2}.$$

By a theorem of Hoeffding [3]

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(EZ_{Ni}^{k})^{2} = EX^{2k} \qquad \text{for } k \geqq 1. \tag{2}$$

Hence $\lim_{N\to\infty} U_N = 0$, which proves the theorem.

3. Applications.

EXAMPLE 1. Consider the test studied by Hotelling and Pabst [5] based on the statistic $S_N' = \sum_{i=1}^{N} iR_i$. This statistic was shown to be asymptotically normal in [5]. If we set $a_{Ni} = (i - (N+1)/2)/N(N^2-1)/12$ and $b_{Ni} = i/(N+1) - 1/2$, then it is easy to see that the random variable X which has uniform distribution on the unit interval satisfies the conditions of the theorem with $k = 1$. Hence S_N is asymptotically normal and therefore so is S_N'.

EXAMPLE 2. The statistic $S_N = \sum_{i=1}^{N} a_{Ni}EZ_{Ni}$, where the Z_{Ni} are order statistics from a normal (0, 1) population and the a_{Ni} satisfy certain conditions, has been studied by Hoeffding and others [8] and shown to be asymptotically normal. Our theorem shows S_N to be asymptotically normal not only for the case of normal order statistics but also when the Z_{Ni} are order statistics from any population satisfying conditions (a), (c) and (e). The last will be satisfied, for instance, when

$$a_{Ni} = \begin{cases} \sqrt{n/(mN)} & (i = 1, \cdots, m) \\ -\sqrt{m/(mN)} & (i = m+1, \cdots, m+n), \end{cases} \tag{3}$$

where $m + n = N$ and m and n both approach infinity as N approaches infinity. This type of a_{Ni} is commonly used in the "two-sample problem."

EXAMPLE 3. When $a_{Ni}[\sum_{i=1}^{N}(EZ_{Ni} - \sum_{i=1}^{N} EZ_{Ni}/N)^2]^{\frac{1}{2}} = EZ_{Ni} - \sum_{i=1}^{N} EZ_{Ni}/N$ and $b_{Ni} = EZ_{Ni} - \sum_{i=1}^{N} EZ_{Ni}/N$, this S_N has been studied by Hoeffding [2] for the case of Z_{Ni} from a normal (0, 1) population. In this case he showed S_N to be asymptotically normal. Our theorem shows this is also true when the Z_{Ni} are order statistics from any population satisfying (a) and (c), ($k = 1$), since (e) holds. This is so since $\max_{1\leqq i\leqq N} |a_{Ni}|$ is given for either the index 1 or N. Assume it is N. We have $EZ_{NN}^{j} = N\int_{-\infty}^{\infty} x^{j}F^{N-1}(x)dF(x)$, ($j = 1, 2$), and an easy argument gives that $\lim_{N\to\infty} EZ_{NN}^{j}/N = 0$. This and the fact that $(EZ_{NN})^2 \leqq EZ_{NN}^2$ together with (2) proves the assertion. If the index is 1, the proof is analogous.

EXAMPLE 4. When the a_{Ni} are given by (3) and $b_{Ni} = i/(N+1) - \frac{1}{2}$ the statistic S_N is, for every N, linearly related to the Wilcoxon statistic, further discussed by Mann and Whitney [6], which, as is well known, is asymptotically normal. This is also seen from our theorem for reasons stated in Examples 1 and 2.

EXAMPLE 5. In a thesis by Terry [8], the statistic $m - \sum_{i=1}^{m} EZ_{NR_i}^2$ (where the Z_{Ni} are the order statistics from a normal (0, 1) population) is proposed against the alternative that the X_i are normal with common mean, the first m having one variance, the remaining $M - n$ another. This statistic is linearly related to an S_N where the a_{Ni} are given by (3) and $b_{Ni} = EZ_{Ni}^2 - \sum_{j=1}^{N} EZ_{Nj}^2/N$. No consideration of the asymptotic distribution of this statistic is made in [8]. We see that this S_N is asymptotically normal when the Z_{Ni} are order statistics from any population satisfying (a) and (c).

By way of example of a case not covered by earlier theorems (for instance, see Theorem 4 of [4]) we take $S_N = \sum_{i=1}^{N} a_{Ni} EZ_{NR_i}$ where the Z_{Ni} are order statistics from a normal (0, 1) population and where condition (13) of [4] is not satisfied. We can construct such a case as follows. Let the a_{Ni} be given by (3) but let the integer m be fixed and independent of N. Then condition (13) of [4] says that

$$\left[n\left(\frac{m}{nN}\right)^{r/2} + m\left(\frac{-n}{mN}\right)^{r/2}\right] \frac{\sum_{i=1}^{N} EZ_{Ni}^r/N}{\left[\sum_{i=1}^{N} EZ_{Ni}^2/N\right]^{r/2}} \tag{4}$$

must approach zero as N approaches infinity for $r = 3, 4, \cdots$. From [3] we have that $\sum_{i=1}^{N} EZ_{Ni}^j/N$ has for its limit the jth moment of a normal (0, 1) variable. Hence for even r, (4) does not approach zero. However, we see from our theorem that S_N is asymptotically normal.

REFERENCES

[1] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] W. HOEFFDING, "Most powerful rank order tests of independence," unpublished paper, 1950.

[3] W. HOEFFDING, "On the distribution of the expected values of order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 93–100.

[4] W. HOEFFDING, "A combinatorial central limit theorem," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 558–566.

[5] H. HOTELLING AND M. R. PABST, "Rank correlation and tests involving no assumption of normality," *Ann. Math. Stat.*, Vol. 7 (1936), pp. 29–43.

[6] H. B. MANN AND R. R. WHITNEY, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 50–60.

[7] G. E. NOETHER, "On a theorem by Wald and Wolfowitz," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 455–458.

[8] M. TERRY, "Some rank order tests which are most powerful against specific parametric alternatives," Univeristy of North Carolina Library Unpublished dissertation, 1951.

[9] A. WALD AND J. WOLFOWITZ, "Statistical tests based on permutations of the observations," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 358–372.

NOTE ON THE VARIATION OF MEANS

By Casper Goffman

Wayne University

In a manufactured product, batch to batch variations may appear, and it may be of interest to be able to compare these variations for different runs. The simplest case is that for which there is normal distribution with the same standard deviation for each batch, but where the mean may vary from batch to batch. The question arises regarding what function of the set of means should be taken as a measure of its variation. Thus, if $x_1, x_2, \cdots, x_n$ are independent random variables, all with the same standard deviation, say $\sigma = 1$, and means $\mu_1, \mu_2, \cdots, \mu_n$, the question is what function $f(\mu_1, \mu_2, \cdots, \mu_n)$ should be taken to measure the variation of the means. We find, in this note, that if $f(\mu_1, \mu_2, \cdots, \mu_n)$ is subjected to four conditions, three of which seem quite natural and the fourth of which, although perhaps not so natural, has a certain appeal, then $f(\mu_1, \mu_2, \cdots, \mu_n) = F(V)$, where V is the sum of squares $\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2$, $\bar{\mu} = \sum_{i=1}^{n} \mu_i/n$. The properties we have in mind are:

(i) $f(\mu_1, \mu_2, \cdots, \mu_n)$ is continuous, nonnegative, and is equal to zero if and only if $\mu_1 = \mu_2 = \cdots = \mu_n$.

(ii) For every $\epsilon > 0$, there is a $\delta > 0$, such that whenever $f(\mu_1, \mu_2, \cdots, \mu_n) < \delta$ then $|\mu_i - \mu_j| < \epsilon$ for every $i, j = 1, 2, \cdots, n$.

(iii) For every $\mu_1, \mu_2, \cdots, \mu_n$ and every h, $f(\mu_1 + h, \cdots, \mu_n + h) = f(\mu_1, \cdots, \mu_n)$.

(iv) If $x_1, x_2, \cdots, x_n; x_1', x_2', \cdots, x_n'$ are normally distributed with standard deviation $\sigma = 1$ and means $\mu_1, \cdots, \mu_n; \mu_1', \cdots, \mu_n'$, respectively, and if $f(\mu_1, \cdots, \mu_n) = f(\mu_1', \cdots, \mu_n')$, then the random variables $u = f(x_1, \cdots, x_n)$ and $v = f(x_1', \cdots, x_n')$ have the same distribution function.

Condition (iv) says that the distribution of the estimate of the variation of means, obtained from samples, depends only upon the measures of the variation of the means, (assuming standard deviation 1) and upon no other aspect of the set of means.

In this connection, we note that the distribution of the sum of squares of n independent variables with means $a_1, a_2, \cdots, a_n$ depends only on the variance of the means, as does the power function [1] of the analysis of variance test.

Theorem. *If $f(\mu_1, \cdots, \mu_n)$ has properties* (i)–(iv), *there is a continuous $F(x)$ such that $F(V) = f(\mu_1, \cdots, \mu_n)$, where $V = \sum_{i=1}^{n} (\bar{\mu} - \mu_i)^2$, $\bar{\mu} = \sum_{i=1}^{n} \mu_i/n$.*

Proof. Let $x_1, \cdots, x_n; x_1', \cdots, x_n'$ be normal, with means $\mu_1, \cdots, \mu_n; \mu_1', \cdots, \mu_n'$ and standard deviation $\sigma = 1$. Suppose

$$\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2 \neq \sum_{i=1}^{n} (\mu_i' - \bar{\mu}')^2. \tag{1}$$

Received 7/10/52.

By property (iii), we may suppose that $\sum_{i=1}^{n} \mu_i = \sum_{i=1}^{n} \mu_i' = 0$. Then (1) becomes

$$\sum_{i=1}^{n} \mu_i^2 \neq \sum_{i=1}^{n} \mu_i'^2. \tag{2}$$

Let $\alpha = [\sum_{i=1}^{n} \mu_i^2]^{\frac{1}{2}}$, $\beta = [\sum_{i=1}^{n} \mu_i'^2]^{\frac{1}{2}}$ and suppose $\beta > \alpha$. Let $\epsilon = (\beta - \alpha)/(2n + 1)$. By property (ii), there is a $\delta > 0$ such that if $f(x_1, \cdots, x_n) < \delta$ then $| x_i - x_j | < \sqrt{n}\,\epsilon$ for all $i, j = 1, 2, \cdots, n$. Now, let E be the set of points, $(x_1, \cdots, x_n)$, for which $f(x_1, \cdots, x_n) < \delta$, and let $E_0 \subset E$ be those points of E for which $\sum_{i=1}^{n} x_i = 0$. Then

$$\begin{aligned} P(u < \delta) &= (\tfrac{1}{2}\pi)^{n/2} \int_E \cdots \int e^{-\frac{1}{2}\sum_{i=1}^{n} (\mu_i - x_i)^2}\, dx_1 \cdots, dx_n \\ &\geqq (\tfrac{1}{2}\pi)^{n/2} \mid E_0 \mid \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(\alpha + n\epsilon)^2 + x^2]}\, dx, \end{aligned} \tag{3}$$

$$\begin{aligned} P(v < \delta) &= (\tfrac{1}{2}\pi)^{n/2} \int_E \cdots \int e^{-\frac{1}{2}\sum_{i=1}^{n} (\mu_i' - x_i')^2}\, dx_1, \cdots, dx_n \\ &\leqq (\tfrac{1}{2})^{n/2} \mid E_0 \mid \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(\beta - n\epsilon)^2 + x^2]}\, dx, \end{aligned} \tag{4}$$

where $| E_0 |$ is the $n - 1$ dimensional measure of E_0.

That expressions (3) and (4) hold may be shown as follows: E is the cylinder whose axis is the line $x_1 = x_2 = \cdots = x_n$. The n-tuple integrals may be evaluated by letting x be measured along this axis and by integrating, for each x, over the hyperplane normal to the axis at x, and then by integrating with respect to x. It follows that $P(u < \delta) \geqq (\frac{1}{2}\pi)^{n/2} \mid E_0 \mid \int_{-\infty}^{\infty} \varphi(x)\, dx$ and $P(v < \delta) \leqq (\frac{1}{2}\pi)^{n/2} \mid E_0 \mid \int_{-\infty}^{\infty} \psi(x)\, dx$, where $\varphi(x) \leqq e^{-\frac{1}{2}\sum_{i=1}^{n} (\mu_i - x_i)^2}$ and $\psi(x) \geqq e^{-\frac{1}{2}\sum_{i=1}^{n} (\mu' i - x' i)^2}$ for all $(x_1, \cdots, x_n)$ and $(x_1', \cdots, x_n')$ in E whose projections on the axis $x_1 = x_2 = \cdots = x_n$ fall at x. Now, for every such $(x_1, \cdots, x_n)$, $\sum_{i=1}^{n} (\mu_i - x_i)^2 \leqq x^2 + (\alpha + n\epsilon)^2$, since the vector whose components are $\mu_i - x_i$, $i = 1, \cdots, n$, has x as one orthogonal component, the other of which is not greater than the sum of the distances of $(\mu_1, \cdots, \mu_n)$ and $(x_1, \cdots, x_n)$ from the line $x_1 = x_2 = \cdots = x_n$; but this is readily seen not to to exceed $\alpha + n\epsilon$. Similarly, $\sum_{i=1}^{n} (\mu_i' - x_i')^2 \geqq x^2 + (\beta - n\epsilon)^2$. Accordingly, we may take $\varphi(x) = e^{-\frac{1}{2}[(\alpha + n\epsilon)^2 + x^2]}$ and $\psi(x) = e^{-\frac{1}{2}[(\beta - n\epsilon)^2 + x^2]}$. Moreover,

$$| E_0 | > 0, \tag{5}$$

for, since $f(x_1, \cdots, x_n)$ is continuous, there is a sphere S of radius less than ϵ, containing $(0, 0, \cdots, 0)$, for every point $(x_1, \cdots, x_n)$ of which $f(x_1, \cdots, x_n) < \delta$. The set $S_0 \subset S$, of points $(x_1, \cdots, x_n) \varepsilon S$ for which $\sum_{i=1}^{n} x_i = 0$, is a subset of E_0 of positive $n - 1$ dimensional measure. But $\epsilon = (\beta - \alpha)/(2n + 1)$ implies

$\beta - n\epsilon > \alpha + n\epsilon$. It then follows by (3), (4), and (5) that if $\mu_1, \cdots, \mu_n$; $\mu_1', \cdots, \mu_n'$ satisfies (1), then $P(u < \delta) > P(v < \delta)$. Hence, by property 4, $f(\mu_1, \cdots, \mu_n) \neq f(\mu_1', \cdots, \mu_n')$.

On the other hand, let $\mu_1, \cdots, \mu_n$; $\mu_1', \cdots, \mu_n'$ be such that

$$V = \sum_{i=1}^{n} (\bar{\mu} - \mu_i)^2 = \sum_{i=1}^{n} (\mu_1' - \bar{\mu}')^2. \tag{6}$$

Suppose $a = f(\mu_1, \cdots, \mu_n)$, $b = f(\mu_1', \cdots, \mu_n')$, and that $a \neq b$. Let C_1 and C_2 be continuous curves joining $(\mu_1, \cdots, \mu_n)$ to $(\mu_1', \cdots, \mu_n')$ such that for every $(\mu_1^{(1)}, \cdots, \mu_n^{(1)}) \in C_1$, not an end-point, $\sum_{i=1}^{n} (\mu_i^{(1)} - \bar{\mu}^{(1)})^2 < V$, and for every $(\mu_1^{(2)}, \cdots, \mu_n^{(2)}) \in C_2$, not an end-point, $\sum_{i=1}^{n} (\mu_i^{(2)} - \bar{\mu}^{(2)})^2 > V$. Since $f(\mu_1, \cdots, \mu_n)$ is continuous, there are points $(\mu_1^{(1)}, \cdots, \mu_n^{(1)}) \in C_1$ and $(\mu_1^{(2)}, \cdots, \mu_n^{(2)}) \in C_2$ for which

$$f(\mu_1^{(1)}, \cdots, \mu_n^{(1)}) = f(\mu_1^{(2)}, \cdots, \mu_n^{(2)}) = \tfrac{1}{2}(a + b). \tag{7}$$

But (7) contradicts the fact, already established, that $\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2 \neq \sum_{i=1}^{n} (\mu_i' - \bar{\mu}')^2$ implies $f(\mu_1, \cdots, \mu_n) \neq f(\mu_1, \cdots, \mu_n)$. We have now proved that $f(\mu_1, \cdots, \mu_n) = f(\mu_1', \cdots, \mu_n')$ if and only if $\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2 = \sum_{i=1}^{n} (\mu_i' - \bar{\mu}')^2$. But this is simply another way of saying that there is an $F(x)$ such that $F(V) = f(\mu_1, \cdots, \mu_n)$.

Conversely, it is easy to prove that:

If $F(x)$ is continuous, monotonically increasing, and $F(0) = 0$, then $f(\mu_1, \cdots, \mu_n) = F(V)$ has properties (i)–(iv).

REFERENCE

[1] P. C. Tang, "The power function of the analysis of variance test with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 126–157.

ON MILL'S RATIO FOR THE TYPE III POPULATION

By Des Raj[1]

University of Lucknow

1. Introduction and summary. Mills [1], Gordon [2], Birnbaum [3], and the author [4] have studied the ratio of the area of the standardized normal curve from x to ∞ and the ordinate at x. The object of this note is to establish the monotonic character of, and to obtain lower and upper bounds for, the ratio of the ordinate of the standardized Type III curve at x and the area of the curve from x to ∞. This ratio, as shown by Cohen [5] and the author [6], has to be calculated for several values of x when solving approximately the equations involved in the problem of estimating the parameters of Type III populations from truncated samples. It was found by the author that, for large values of

Received 2/28/52, revised 12/31/52.

[1] Now at the Indian Statistical Institute.

x, when the ordinates and areas are small, either this ratio cannot be obtained from existing tables prepared by Salvosa [7] or that very few significant digits are available for its calculation. It was thus found desirable to obtain lower and upper bounds which could satisfactorily locate this ratio. The monotonic behavior of this ratio and the inequalities obtained may also prove useful in checking the accuracy of the tables in [7], and in studying the nature of the tail of the chi square distribution.

2. Derivations. The standardized Type III population is given by

$$Cf(x)\,dx, \qquad -2/\alpha_3 \leqq x \leqq \infty, \qquad 0 \leqq \alpha_3 \leqq 2, \tag{1}$$

where

$$f(x) = \left\{1 + \frac{\alpha_3}{2}x\right\}^{(4/\alpha_3^2)-1} e^{-(2/\alpha_3)x}, \tag{2}$$

and

$$C = (4/\alpha_3^2)^{(4/\alpha_3^2)-\frac{1}{2}} e^{-4/\alpha_3^2} [\Gamma(4/\alpha_3^2)]^{-1}.$$

We define

$$\mu(x) = f(x)[G(x)]^{-1}, \tag{3}$$

where

$$G(x) = \int_x^\infty f(t)\,dt. \tag{4}$$

We have

$$\mu(-2/\alpha_3) = 0, \qquad \mu(\infty) = 2/\alpha_3,$$

and

$$\frac{d}{dx}\mu(x) = \mu(x)[G(x)]^{-1}\phi_1(x), \tag{5}$$

where

$$\phi_1(x) = f(x) - \left(x + \frac{\alpha_3}{2}\right)\left(1 + \frac{\alpha_3}{2}x\right)^{-1} G(x). \tag{6}$$

Since

$$\phi_1(-2/\alpha_3) = \infty, \qquad \phi_1(\infty) = 0,$$

and

$$\frac{d}{dx}\phi_1(x) = -\left(1 + \frac{\alpha_3}{2}x\right)^{-2}\left(1 - \frac{\alpha_3^2}{4}\right)G(x) \leqq 0,$$

it follows that $\mu(x)$ is monotonically increasing and that

$$\mu(x) \geqq \mu_1(x) = \left(x + \frac{\alpha_3}{2}\right)\left(1 + \frac{\alpha_3}{2}x\right)^{-1}. \tag{7}$$

Again, considering

$$\phi_2(x) = -\left(x + \frac{\alpha_3}{2}\right)\left(1 + \frac{\alpha_3}{2}x\right)^{-1} f(x) + \left[\left(x + \frac{\alpha_3}{2}\right)^2 + \frac{\alpha_3}{2}\left(\frac{2}{\alpha_3} - \frac{\alpha_3}{2}\right)\right]\left(1 + \frac{\alpha_3}{2}x\right)^{-2} G(x), \tag{8}$$

we have

$$\phi_2(-2/\alpha_3) = \infty, \qquad \phi_2(\infty) = 0,$$

TABLE I

Values of $\mu(x)$, $\mu_1(x)$, $\mu_2(x)$, and $\mu_3(x)$

$\alpha_3 = 1.0$

x	$\mu_1(x)$	$\mu(x)$	$\mu_3(x)$	$\mu_2(x)$
−.50	0.000	0.692	0.869	∞
.00	0.500	0.901	1.000	2.000
.50	0.800	1.059	1.117	1.400
1.00	1.000	1.180	1.215	1.333
1.50	1.143	1.275	1.298	1.357
2.00	1.250	1.351	1.366	1.400
2.50	1.330	1.413	1.423	1.444
3.00	1.400	1.464	1.472	1.486
3.50	1.455	1.507	1.513	1.523
4.00	1.500	1.544	1.549	1.556

and

$$\frac{d}{dx}\phi_2(x) = -\alpha_3\left(\frac{2}{\alpha_3} - \frac{\alpha_3}{2}\right)\left(1 + \frac{\alpha_3}{2}x\right)^{-2}[G(x)]^{-1}\left[\mu(x) - x\left(1 + \frac{\alpha_3}{2}x\right)^{-1}\right] \leqq 0,$$

so that

$$\mu(x) \leqq \mu_2 x = \frac{\left(x + \frac{\alpha_3}{2}\right)^2\left(1 + \frac{\alpha_3}{2}x\right)^{-2} + \frac{\alpha_3}{2}\left(\frac{2}{\alpha_3} - \frac{\alpha_3}{2}\right)\left(1 + \frac{\alpha_3}{2}x\right)^{-2}}{\left(x + \frac{\alpha_3}{2}\right)\left(1 + \frac{\alpha_3}{2}x\right)^{-1}}, \tag{9}$$

$$x \geqq -\frac{\alpha_3}{2}.$$

Combining (7) and (9) we have the inequalities

$$\mu_1(x) \leqq \mu(x) \leqq \mu_2(x).$$

A better estimate for the upper inequality can be obtained from Jensen's inequality.

$$(10) \qquad \phi\left[\int_a^b tg(t)\,dt \Big/ \int_a^b g(t)\,dt\right] \leqq \int_a^b \phi(t)g(t)\,dt \Big/ \int_a^b g(t)\,dt,$$

where $\phi(t)$ is convex and $g(t) \geqq 0$ in (a, b). Setting $a = x$, $b = \infty$, $\phi(t) = (1 + (\alpha_3 t)/2)(t + \alpha_3/2)^{-1}$, and $g(t) = (t + \alpha_3/2)(1 + (\alpha_3 t)/2)^{(4/\alpha_3^2)-2} e^{-2t/\alpha_3}$, in (10), we have

$$(11) \qquad -\left(1 + \frac{\alpha_3}{2}x\right)\mu^2(x) + x\mu(x) + 1 \geqq 0,$$

from which it follows that

$$(12) \qquad \mu(x) \leqq \mu_3(x) = 2\left[-x + \sqrt{x^2 + 4\left(1 + \frac{\alpha_3}{2}x\right)}\right]^{-1}.$$

As a check on our results, by putting $\alpha_3 = 0$ in (7), (9) and (12), we obtain inequalities given in [2], [3] and [4]. Incidently, the function

$$\phi_3(x) = \left(1 + \frac{\alpha_3}{2}x\right)\mu(x) - x,$$

used by Cohen [5] can be shown to be monotonically decreasing. For,

$$\phi_3(-2/\alpha_3) = 2/\alpha_3, \qquad \phi_3(\infty) = 0,$$

and

$$\frac{d}{dx}\phi_3(x) = \left(1 + \frac{\alpha_3}{2}x\right)\mu^2(x) - x\mu(x) - 1 \leqq 0 \quad \text{from (11).}$$

The closeness with which these inequalities can locate $\mu(x)$ is illustrated by Table I, where $\mu(x)$ is calculated from the tables in [7].

REFERENCES

[1] J. P. Mills, "Table of the ratio: area to bounding ordinate, for any portion of the normal curve," *Biometrika*, Vol. 18 (1926), pp. 395–400.

[2] R. D. Gordon, "Values of Mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 364–366.

[3] Z. W. Birnbaum, "An inequality for Mill's ratio," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 245–246.

[4] Des Raj, "On estimating the parameters of normal populations from singly truncated samples," *Ganita*, Vol. 3 (1952), pp. 41–57.

[5] A. C. Cohen, "Estimating parameters of Pearson Type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 45 (1950), pp. 411–423.

[6] Des Raj, "On estimating the parameters of Type III populations from truncated samples," *J. Amer. Stat. Assn.*, to be published.

[7] L. R. Salvosa, "Tables of Pearson's Type III function," *Ann. Math. Stat.*, Vol. 1 (1930), appended.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Om P. Aggarwal, formerly Research Assistant at Stanford, received his Ph.D. degree in statistics from Stanford University and is now Assistant Professor in the Department of Mathematics, University of Washington, Seattle.

Dr. P. H. Anderson has accepted a position as a price economist with the Office of Price Stabilization, Washington, D. C.

Mr. George F. Cramer has accepted a position as Staff Scientist with the recently established Computing Center of the Engineering Research Associates Division of Remington Rand Inc. at Arlington, Virginia.

Dr. Harald Cramér, Rector of the University of Stockholm, has accepted a Visiting Professorship in the Department of Mathematics, University of California, Berkeley, for the Fall Semester, 1953, and will be in the Statistical Laboratory beginning the middle of July.

Mr. Richard DeLancie has been released from active duty as a Lieutenant Commander in the Naval Reserves to accept a civilian position as an Operations Analyst with Headquarters, Western Air Defense Force, Hamilton Air Force Base, California.

Dr. Francis W. Dresch, formerly Director of Computation and Ballistics, U. S. Naval Proving Ground, Dahlgren, Virginia, is working in the Economics Division, Operations Research, and acting as consultant on statistics, mathematics and computing machinery at Stanford Research Institute, Stanford, California.

Dr. Meyer Dwass, formerly with the Office of the Assistant Director for Statistical Standards, U. S. Bureau of the Census, has joined the staff of the Department of Mathematics, Northwestern University as an Assistant Professor.

Professor Evelyn Fix was granted a leave of absence January 5 to March 31, 1953 from the Statistical Laboratory, University of California, Berkeley to teach at the Sampling Demonstration Center for the Government of Thailand under the auspices of the Food and Agriculture Organization of the United Nations.

Professor H. M. Gehman, Chairman of the Department of Mathematics of the University of Buffalo, has been serving as Acting Dean of the Graduate School of Arts and Sciences during the first semester of 1952–53.

Dr. T. N. E. Greville of the National Office of Vital Statistics, Public Health Service, has been appointed a technical adviser in Vital and Health Statistics at the Institute of Inter-American Affairs, Rio de Janeiro, Brazil.

Mr. Charles H. Hubbell has completed a tour of active duty with Project *Scoop*, HQ, USAF and is now a research assistant in the Decision Processes Research Project of the Ford Foundation at the University of Michigan.

Dr. Gopinath Kallianpur has returned to India after spending the last academic year at the Institute for Advanced Study, Princeton, New Jersey.

Mr. Wharton F. Keppler, formerly Statistical Analysis Director for M. & R. Dietetic Laboratories, Inc., has been since March, 1952, Staff Analytical Statistician (Engineering) to the Quality Control Directorate, Central Air Procurement District, USAF, Detroit, Michigan.

Mr. Nathan Keyfitz has been granted a year's leave of absence by the Dominion Bureau of Statistics in Ottawa to serve as statistical consultant in Indonesia under the United Nations Technical Assistance Program.

Mr. Gunnar Kulldorf from the University of Lund, Sweden, was awarded a Fellowship from the Sweden-America Foundation and has spent the academic year 1952–53 in the Statistical Laboratory, University of California, Berkeley, and in the Department of Mathematical Statistics, University of North Carolina, Chapel Hill.

Dr. Charles R. Langmuir has been granted a leave from Syracuse University, and has been elected a member and director of the Educational Research Corporation and appointed Research Associate.

Mr. R. B. Murphy, formerly at Carnegie Institute of Technology, is now associated with the Bell Telephone Laboratories, Inc.

Dr. Donald B. Owen, formerly Instructor in mathematics at the University of Washington, has joined the staff of the Mathematics Department at Purdue University as an Assistant Professor and Consultant in the Statistical Laboratory.

Dr. Stefan Peters has accepted the position of actuary with Morss and Seal, Consulting Actuaries, New York.

Mr. Sixto Rios, Catedrático of Mathematical Statistics, University of Madrid, has been appointed as Director of the Statistical School, University of Madrid.

Dr. Murray Rosenblatt, Assistant Professor of Statistics at the University of Chicago, will visit the Institute of Mathematical Statistics in Stockholm from April to December, 1953. He will continue research with Ulf Grenander on time series analysis initiated when the latter spent the year 1951–52 as Visiting Assistant Professor at Chicago.

Professor Henry Scheffé, Department of Mathematical Statistics, Columbia University, has accepted the position of Professor and Assistant Director of the Statistical Laboratory, University of California, Berkeley, beginning with the Fall term 1953.

Dr. R. W. Shephard of the Rand Corporation has accepted a position with the Sandia Corporation, Albuquerque, New Mexico.

Professor Jack Silber of Roosevelt College has returned from a tour of duty as Operations Analyst with the Fifth Air Force in Korea.

Mr. Oliver A. Shaw, recalled to active duty with the Air Force in 1951, is now stationed at Mather AFB, California. Major Shaw was teaching mathematics at the University of Mississippi at the time he was recalled to active duty.

Mr. John R. Sullivan has returned to his position as Assistant Professor of Mathematics at Clemson College, South Carolina, after a two-year leave of absence for graduate study in the Department of Mathematical Statistics at the University of North Carolina.

Summer Statistical Seminar at the University of Connecticut

The University of Connecticut will hold its fourth annual Summer Seminar in Statistics from August 10 through 28, 1953. The general plan provides for one or two seminar sessions daily and a clinic on the treatment of problems in application.

The subjects for discussion together with their organizers are: August 10–14, "Statistical Methodology in Physics," Dr. E. W. Pike with Dr. Churchill Eisenhart and Dr. E. P. King; August 17–21, "Statistics in Biometry and Medicine," Professor G. Beall with Dr. I. Bross and Dr. D. Mainland; August 24–28, (i) "ASA Handbook," Professor F. Mosteller, (ii) "Performance and Reliability of Complex Mechanical Assemblies," Professor G. H. Shortley.

Persons interested are invited to attend any or all sessions. For more detailed information, write the Secretary of the Seminar, Professor Geoffrey Beall, Statistical Laboratory, University of Connecticut, Storrs, Connecticut.

Doctoral Dissertations in Statistics, 1952

Listed below are the doctorates conferred during the year 1952 in the United States and Canada for which the dissertations were written on topics in statistics (or for a degree in statistics). The university, month in which degree was conferred, major subject, minor subject, and the title of the dissertation are given in each case if available.

A. G. Anderson, Michigan, February, major in statistics, "The Prediction of Quantitative Characteristics in Polygenic Systems."

J. B. Bartoo, State University of Iowa, June, major in mathematics, "Certain Theorems on Order Statistics."

C. A. Bennett, Michigan, June, major in statistics, "Asymptotic Properties of Ideal Linear Estimators."

W. H. Clatworthy, North Carolina, August, major in mathematical statistics, "Partially Balanced Incomplete Block Designs with $r < k$."

E. L. Cox, North Carolina, major in experimental statistics, "On Estimating Size of Biological Populations."

M. Dwass, North Carolina, August, major in mathematical statistics, "On the Large Sample Power of Certain Rank Order Tests."

F. E. Freund, Pittsburgh, August, "Some Methods of Estimating Prior Probabilities from Heterogeneous Populations."

S. G. Ghurye, North Carolina, August, major in mathematical statistics, "Some Problems in the Theory of Stochastic Difference Equations."

F. A. Graybill, Iowa State College, June, major in statistics, "On Quadratic Estimates of Variance Components."

W. C. Guenther, Washington, December, major in mathematical statistics, "On Testing Whether or Not a Given Percentile of One Distribution is Less Than or Equal to a Given Percentile of Another Distribution."

K. D. C. Haley, Stanford, major in statistics, minor in mathematics, "Estimation of the Dosage Mortality Relationship When the Dose is Subject to Error."

A. T. James, Princeton, October, major in statistics, "Group Methods in Normal Multivariate Distribution Theory."

T. A. Jeeves, California (Berkeley), June, major in statistics, "Identifiability and Almost-Sure Estimability of Linear Structures in n-Dimensions."

J. C. Kiefer, Columbia, May, major in mathematical statistics, "Contributions to the Theory of Games and Statistical Decision Functions."

L. M. Le Cam, California (Berkeley), June, major in statistics, "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates."

M. R. Mickey, Jr., Iowa State College, July, major in statistics, "An Application of Sequential Tests to a Problem of Quality Control."

W. J. Moonan, Minnesota, December, major in statistics, "The Generalization of the Principles of Some Modern Experimental Designs for Psychological and Educational Research."

H. Raiffa, Michigan, February, major in statistics, "Arbitration Systems for Generalized Two-Person Games."

N. Rudy, Chicago, August, major in statistics, "Some Problems in the Economics of Industrial Sampling Inspection."

A. R. Sen, North Carolina, major in experimental statistics, "Further Developments of the Theory and Application of the Selection of Primary Sampling Units, with Special Reference to the North Carolina Agricultural Population."

K. W. Smillie, University of Toronto, June, major in statistics, "A Mathematical Treatment of Certain Movements of Fish—An Application of the Theory of Markov Processes."

R. F. Tate, California (Berkeley), June, major in statistics, "Contributions to the Theory of Random Numbers of Random Variables."

D. J. Thompson, Iowa State College, August, major in statistics, "A Theory of Sampling Finite Universes with Arbitrary Probabilities."

C. K. Tsao, Oregon, June, major in mathematical statistics, "A General Class of Discrete Distributions and Mixtures of Distributions."

G. S. Watson, North Carolina, major in experimental statistics, "Serial Correlation in Regression Analysis."

R. E. Wheeler, Kentucky, June, "A Variable Probability Distribution Function."

P. Whidden, Carnegie Institute of Technology, June, major in mathematics, "A Criterion for Measuring Closeness of Probability Distributions."

The following names were omitted from the 1951 list of doctoral dissertations in statistics.

R. W. Allen, St. Louis, June, 1951, "Compound Statistical Distribution Functions."

W. W. Jacobs, George Washington University, May, 1951, major in mathematical statistics, "Random Matrices."

New Members

The following persons have been elected to membership in the Institute

(November 27, 1952 to February 18, 1953)

BARTOO, JAMES B., Ph.D. (State Univ. Iowa), Assistant Professor, Department of Mathematics, Pennsylvania State College, State College, Pennsylvania.

BERSHAD, MAX A., B.S. (College of the City of New York), Statistician, Statistical Research Section, Bureau of the Census, *1661 Ft. Du Pont St., S.E., Washington 20, D. C.*

BINGHAM, RICHARD STEPHEN JR., B.S. (Carnegie Inst. of Tech.), Quality Control Supervisor, Atlas Powder Company, Volunteer Ordnance Works, *1413 Wright Street, Chattanooga, Tennessee.*

CLARKE, A. BRUCE, Ph.D. (Brown Univ.), Instructor, Department of Mathematics, University of Michigan, Ann Arbor, Michigan.

DEEKS, HERBERT W. G., M.S. (Univ. of London), Statistician to the Directorate General of Works in the Ministry of Works, London, *Eversfield Red Hill, Chislehurst Kent, England.*

EICH, EDWARD D., B.S. (Mass. Inst. of Tech.), Electrical Engineer, Anaconda Wire & Cable Company, *Anaconda Wire & Cable Company, Hastings-on-Hudson, New York.*

EKBLOM, STAFFAN, F.K. (Univ. of Stockholm), Research Assistant, *Norrtullsgatan 16, Stockholm, Sweden.*

EKLIND, JAN-ROBERT, F.K. (Univ. of Stockholm), Instructor in Mathematical Statistics, University of Stockholm, *Norrtullsgatan 16, Stockholm, Sweden.*

FEDERSPIEL, CHARLES F., A.M. (Univ. of Michigan), Statistician, Communicable Disease Center, U. S. Public Health Service, *169 Eighth St. N.E., Atlanta 5, Georgia.*

GILBERT, JOHN P., B.A. (St. John's College), Supervisor of Computers, Statistical Research Center, University of Chicago, *5719 Dorchester Avenue, Chicago 37, Illinois.*

HOPPER, GRACE M., Ph.D. (Yale Univ.), Systems Engineer, Remington Rand Inc., *Walnut Park Plaza, Walnut at 63rd St., Philadelphia 39, Pennsylvania.*

HOY, WALTER W., M.A. (Ohio State Univ.), Graduate Student, Ohio State University, *55½ E. Norwich Avenue, Columbus 1, Ohio.*

HUSEIN, HASAN M., Ph.D. (Leeds Univ.), Professor of Statistics, Faculty of Commerce, Fouad University, Cairo, *28, Mobtadayan Street, Cairo, Egypt.*

KEMPERMAN, JOHAN H. B., Ph.D. (Univ. of Amsterdam), Visiting Professor, Department of Mathematics, Purdue University, West Lafayette, Indiana.

MCQUAID, GERTRUDE A., A.B. (Hunter College), Graduate Student, New York University, *3034 Grand Concourse, New York 58, New York.*

MURPHY, JOHN E., M.A. (Columbia Univ.), Statistical Analyst, Bristol-Myers Products Division, *c/o Bristol-Myers Products Division, 630 Fifth Avenue, New York 20, New York.*

PENG, KAN-CHEN, M.S. (Univ. of Michigan), Statistical Quality Control Engineer, Parke, Davis & Company, *823 Seward, Detroit, Michigan.*

PILLAI, K. C. SREEDHARAN, M.S. (Univ. of Travancore), Research Associate in Statistics and Graduate Student, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

PRICE, BRUCE P., M.M. (Cincinnati Conservatory of Music), Research Analyst, Engineering Department, Analytical Mechanics Section, Southwest Research Institute, *242 Future, San Antonio 2, Texas.*

RICHARDS, ROBERT G., M.A. (Univ. of Calif., Berkeley), Mathematician, Department of the Army at Redstone Arsenal, *Y. M. C. A., Huntsville, Alabama.*

ROBINSON, L. V., Ph.D. (Harvard Univ.), Mathematician, Wright-Patterson Air Base, *373 West First Street, Dayton, Ohio.*

STACEY, ALEC G., Supervisor, Special Surveys Division, Regional Office Dominion Bureau of Statistics, St. John's, *75 St. Clare Avenue, St. John's, Newfoundland, Canada.*

WORTHAM, A. W., M.S. (Oklahoma Agric. and Mech. College), Senior Project Analytical Engineer, Chance Vought Aircraft, *Chance Vought Aircraft, Structures Section, Dallas, Texas.*

PUBLICATIONS RECEIVED

CLARK, C. E., *An Introduction to Statistics*, John Wiley and Sons, Inc., New York, 1953, x + 266 pp., $4.25.

COCHRAN, W. G., *Sampling Techniques*, John Wiley and Sons, Inc., New York, 1953, $6.50.

DOOB, J. L., *Stochastic Processes*, John Wiley and Sons, Inc., New York, 1953, vii + 654 pp., $10.00.

FINNEY, D. J., *An Introduction to Statistical Science in Agriculture*, John Wiley and Sons, Inc., New York, 1953, 179 pp., $3.75.

HARTREE, D. R., *Numerical Analysis*, Oxford University Press, New York, 1953, xiv + 287 pp., $6.00.

Recenseamento Geral do Brasil (1° de Setembro de 1940) Censo Demografico and Censos Economicos, Servico Grafico de Instituto Brasileiro de Geografia e Estatistica, Rio de Janeiro, 1950. (7 volumes in addition to those listed in March and September 1952 and March 1953.)

Recenseamento Geral do Brasil (1.° de Setembro de 1940) Censo Demografico e Censos Economicos, Servico Grafico de Instituto Brasileiro de Geografia e Estatistica, Rio de Janeiro, 1951.

SHEPHARD, R. W., *Cost and Production Functions*, Princeton University Press, Princeton 1953, vii + 104 pp., $2.00.

THRON, W. J., *Introduction to the Theory of Functions of a Complex Variable*, John Wiley and Sons, Inc., New York, 1953, ix + 230 pp., $6.50.

TINBERGEN, J., *On the Theory of Economic Policy*, North-Holland Publishing Co., Amsterdam, 1953, iii + 78 pp., $1.80.

VINCI, FELICE, *Interventi sul nuovo assetto Europeo, sull'accertamento della miseria, sulle ricerche econometriche e sull'illusione finanziaria*, (Istituto di Scienze Economiche e Statistiche, Quaderni XVI), Milan, 1952, 22 pp.

ECONOMETRICA

Journal of the Econometric Society

Contents of Vol. 21, No. 2 - April, 1953

Published Quarterly **Subscription rates available on request**

The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics.

Subscriptions to *Econometrica* and inquiries about the work of the Society and the procedure in applying for membership should be addressed to Rossen L. Cardwell, Acting Secretary, The Econometric Society, The University of Chicago, Chicago 37, Illinois, U. S. A.

BIOMETRIKA

A Journal for the Statistical Study of Biological Problems

Volume 40 **Contents** **Parts 1 and 2, June 1953**

The subscription price, *payable in advance*, is 45s. inland, 54s. export (per volume including postage). Cheques should be drawn to Biometrika and sent to "The Secretary, Biometrika Office, Department of Statistics, University College, London, W.C. 1." All foreign cheques must be in sterling and drawn on a bank having a London agency.